Towards a Unified Model of Occupants' Behaviour and Comfort for Building Energy Simulation

THÈSE N^{O} 4587 (2010)

PRÉSENTÉE LE 5 FÉVRIER 2010

À LA FACULTÉ ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE D'ÉNERGIE SOLAIRE ET PHYSIQUE DU BÂTIMENT
PROGRAMME DOCTORAL EN ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Frédéric HALDI

acceptée sur proposition du jury:

Prof. A. Schleiss, président du jury
Dr D. Robinson, directeur de thèse
Prof. M. Bierlaire, rapporteur
Prof. M. A. Humphreys, rapporteur
Prof. C. Reinhart, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2010

Abstract

The building sector alone accounts for around half of the energy consumed in Switzerland and most other developed countries, with associated adverse environmental consequences, and there is a great potential for savings in this sector. For this reason, the development of efficient solutions for predicting and optimising the energy and environmental performance of buildings is clear.

Dynamic building thermal simulation programs are increasingly used for this purpose. However, some key processes are still not taken into account by these tools, leading to potentially significant errors. Most noteworthy is the influence of buildings' occupants, whose actions such as the use of windows and shading devices have an important impact on the indoor environment and the overall energy performance of a building. Furthermore, occupants' environmental comfort is the central underlying concept influencing actions on building controls; but the intrinsic interaction between these notions is not well known.

This thesis develops adequate models for the prediction of occupants' actions that have an impact on building performance and further proposes an innovative global formulation of the link between environmental comfort, human adaptive actions in the built environment and their feedback in terms of satisfaction and acceptability. Furthermore, detailed integration procedures of these methods into building and urban simulation tools are described.

Based on detailed statistical analyses of eight years of continuous measurements, a model for the prediction of actions on windows performed by office occupants is proposed. It is formulated as an occupancy-dependent Markov chain extended to a continuous-time process for opening durations. The explanatory variables have been carefully selected on the basis of statistical relevance, which are indoor and outdoor temperature, the occurrence of rain, and occupant presence and absence durations. The choice of the specific form of the model is justified by cross-validation and its superior predictive accuracy is determined by comparison with model variants and previously published work.

A similar procedure was carried out for the inference of a model to predict actions on shading devices. Its formulation is also based on rigorously selected predictors used as inputs to an occupancy-dependent Markov chain expressing action probabilities. The model has also been extended to predict the choice of shaded fraction. Once again simulations of model variants support the choice of the final model.

Using results of a long-term survey of building occupants, we evaluate the accuracy of currently accepted models for thermal comfort prediction and identify clear weaknesses. We go on to propose a probabilistic formulation for the distribution of thermal sensation and for the occurrence of the state of thermal comfort and extend this to visual comfort. The result is a simple and accurate definition of comfort probability and its variations amongst individuals.

We have also analysed variables which influence occupants' comfort temperature. This

has enabled us to assign weights to the key variables influencing comfort temperature: adaptation, acclimatisation and individuality. We also consider the feedback of actions on comfort and numerically estimate "adaptive increments to comfort temperature". This results in a proposed formulation for a new adaptive model for thermal comfort, for general application in buildings with variable degrees of adaptation available to occupants.

The link between thermal and visual comfort with actions on windows and shading devices is also studied and formulated as a single unified concept linked by human action inertia whose properties are discussed.

Finally, new modelling approaches have been developed for the prediction of adaptations of personal characteristics such as clothing and metabolic activity, an assessment of the very limited degree of interaction between thermal, olfactory and visual comfort and finally an analysis of factors influencing perceived productivity in office environments, in which hot conditions are shown to cause a decrease of the order of 10% compared to relatively cooler conditions.

Keywords: Building simulation, Behavioural modelling, Agent-based model, Field survey, Windows, Shading devices, Thermal comfort, Visual comfort, Clothing, Metabolic activity, Productivity

Résumé

Le secteur du bâtiment absorbe à lui seul près de la moitié de l'énergie consommée en Suisse ainsi que dans la plupart des autres pays développés. Le potentiel d'économies d'énergie dans ce secteur est donc important. Par conséquent, le développement de méthodes efficaces pour prédire et optimiser la performance énergétique et environnementale des bâtiments est d'un intérêt particulier.

A cette fin, l'utilisation de logiciels de simulation thermique dynamique du bâtiment est de plus en plus répandue. Cependant, ces outils ne prennent toujours pas en compte certains processus importants, conduisant à des erreurs potentiellement significatives.

L'influence des occupants des bâtiments en est la cause principale, du fait que leurs actions, telles que l'usage des fenêtres et des stores, ont un impact direct sur l'environnement intérieur et la performance énergétique globale. En outre, le confort environnemental des occupants constitue le concept central sous-jacent qui détermine les actions sur les contrôles du bâtiment, mais le lien intrinsèque entre ces notions et mal connu.

Cette thèse développe des modèles appropriés pour prédire les actions des occupants de bâtiments ayant un impact sur la performance énergétique. Elle propose également une formulation globale innovante du lien entre confort environnemental et actions adaptatives humaines dans l'environnement construit, ainsi que de leur impact en termes de satisfaction et d'acceptabilité. Elle inclut également une description des procédures détaillées pour l'intégration de ces méthodes dans les outils de simulation thermique à l'échelle des bâtiments et des collectivités urbaines.

A l'aide d'analyses statistiques détaillées de huit ans de mesures continues, nous proposons un modèle pour la prédiction des actions sur les fenêtres effectuées par les occupants d'immeubles de bureaux. Ce modèle est formulé comme une chaîne de Markov dépendant de la présence, étendue à un processus à temps continu pour les durées d'ouverture. Les variables explicatives ont été soigneusement sélectionnées sur la base de leur relevance statistique, incluant les températures intérieure et extérieure, la présence de pluie et les durées d'absence et de présence. Le choix de la forme spécifique du modèle est justifié par validation croisée et sa valeur prédictive supérieure démontrée par comparaison avec des variantes de modélisation ainsi qu'avec des travaux précédemment publiés.

Une procédure similaire a été menée pour l'inférence d'un modèle prédisant les actions sur les stores. Sa formulation est aussi basée sur des variables rigoureusement sélectionnées utilisées dans une chaîne de Markov dépendant de la présence, exprimant les probabilités d'action. Le modèle a également été étendu pour prédire le choix des fractions ombrées. Comme précédemment, des simulations de variantes alternatives de modélisation justifient le choix du modèle final.

A l'aide de résultats issus d'un sondage de longue durée pratiqué sur des occupants de bâtiments, nous évaluons la pertinence des modèles de confort thermique utilisés actuellement et identifions leurs faiblesses. Sur cette base, nous proposons une formulation proba-

biliste pour la distribution de la sensation thermique ainsi que pour la prévalence de l'état de confort thermique et étendons ces concepts au cas du confort visuel. Le résultat consiste en une définition simple et précise de la probabilité de confort et de ses variations parmi les individus.

Nous avons également analysé les variables qui influencent la température de confort des occupants. Ceci permet d'attribuer des pondérations aux variables-clés influençant la température de confort, qui sont l'adaptation, l'acclimatation et les particularités individuelles. Nous considérons également en retour l'influence des actions sur le confort et estimons numériquement des "incréments adaptatifs sur la température de confort". Ceci résulte en une proposition de formulation d'un nouveau modèle adaptatif pour le confort thermique, destiné à l'application généralisée dans des bâtiments offrant divers degrés d'adaptation à leurs occupants.

Le lien entre le confort thermique et visuel d'une part et les actions sur les fenêtres et les stores d'autre part est également étudié et formulé comme un concept unifié dont le lien s'exprime à l'aide de l'inertie d'action, dont les propriétés sont discutées.

Enfin, de nouvelles approches de modélisation ont été développées pour la prédiction des adaptations de caractéristiques personnelles telles que l'habillement et l'activité métabolique, une évaluation du degré d'interaction très limité entre confort thermique, visuel et aéraulique et finalement une analyse des facteurs influençant la productivité perçue dans les environnements de bureau, où il est démontré que des températures élevées causent une baisse de productivité de l'ordre de 10% en comparaison avec des conditions relativement plus fraîches.

Mots-clés: Simulation du bâtiment, Modélisation du comportement, Modèle basé sur les agents, Etude de terrain, Fenêtres, Stores, Confort thermique, Confort visuel, Habillement, Activité métabolique, Productivité

Acknowledgements

I wish to first express my gratitude to Darren Robinson, Claude-Alain Roulet and Jean-Louis Scartezzini for their trust in me and for granting me the possibility to start a thesis in the Solar Energy and Building Physics Laboratory, where I had the privilege of spending a few nice years of fruitful research in a team that joins competence, skills and kindness.

I thank especially my supervisor Darren Robinson, who proposed me the subject of this thesis for which I have developed great interest. All along the preparation of this thesis, he has efficiently supervised this work; contributing ideas, providing sound advice and clearly defining research priorities. This work has also benefited from his careful, rigorous and merciless re-reading.

I express my gratitude to the members of the jury, Anton Schleiss (Laboratory of Hydraulic Constructions, EPFL), Michel Bierlaire (Transport and Mobility Laboratory, EPFL), Michael A. Humphreys (Oxford Brookes University Department of Architecture) and Christoph Reinhart (Harvard University Centre for the Environment) for their careful reading of this thesis, their participation in the examination and their constructive comments that significantly benefited this thesis.

The data which were essential to the models of this thesis would not have been available without the remarkable work of several former collaborators of our laboratory, notably David Lindelöf, Antoine Guillemin and René Altherr, to whom I express my great gratitude. I also thank Laurent Deschamps for his help with the maintenance of our measurement system.

I have also appreciated the technical support of Pierre Loesch and Lise Azzi in the preparation of the measurement campaigns. These measurement campaigns were also greatly facilitated by the collaboration of Christian Trachsel and Robert Sandoz (Department of Urbanism, City Administration of Neuchâtel), Fulvio Faralli (Department of Buildings, Canton of Neuchâtel), Raynald Golay (EOS) and Toni Schroeter (City Administration of Payerne). Many others, who are to numerous to mention by name, have also indirectly helped with these campaigns. Thank you.

My first steps in research have greatly benefitted from the advice of David Lindelöf, who initiated my journey into methods of statistics, data analysis and related digital tools. Several useful discussions with Michel Bierlaire, David Daum, Antoine Guillemin, Michael A. Humphreys, Jérôme Kämpf, Ardeshir Mahdavi, Jessen Page, Claus Pröglhöf and Claude-Alain Roulet have also helped me in developing this work.

I warmly thank the secretaries of our laboratory Suzanne l'Eplattenier, Sylvette Renfer and Barbara Smith for their help in all kinds of matters during my presence at the laboratory and for their lovely company.

A special thank to Lise Azzi, André Kostro and David Daum who have shared my office for several months with a good atmosphere that brightened my every day life as a researcher, even during the final writing-up phase, when I may have been a little taciturn.

My gratitude goes also to all my colleagues of the laboratory who contributed in making it a place of work and study where one simply feels good. I spare a particular thought for my fellows PhD candidates Apiparn Borisuit, Natalia Filchakova, Mario Germano, Martin Joly, Jérôme Kämpf, Philippe Leroux, Friedrich Linhart, Diane Perez, Maria Cristina Munari Probst, Adil Rasheed, Anothai Thanachareonkit, Paola Tosolini and Urs Wilke, with whom I have spent many pleasant moments, numerous morning coffee breaks and joyful fondues.

This research work would not have been possible without the financial support received first by the Swiss National Science Foundation (54^{th} National Research Program, Sustainable Development of the Built Environment) and later by the European Commission (CONCERTO II Project HOLISTIC), which are cordially acknowledged.

My fiends have greatly helped by to keep a sane state of mind during this doctoral research. I thank particularly Olivier Anselmo, Emmanuelle Birraux, Cheryl Chang, Cynthia Dufaux, Christophe Etienne, Julien Ferdinand, Daniel Hairer, Fadi Hamdan, Brigitte Helgen, Khoa Hoàng, Gilles Humbert-Prince, Sara Jones, Aris Khan, Olivier Loutan, Gabrielle Payot, Benjamin Rappaz, Xavier Righetti, Laetitia Stroudinsky, Frank Souiller, Matthieu Sturzenegger, Christophe Tournaire, Julien Vinzio, Pascal Wehrli and Jérôme Wirz for their friendship and the good times we spent together. Particular thanks also go to Christian Theiler, who helped me a lot with relativising the little everyday worries of the researcher.

Words are not enough to express my gratitude to my parents, my dear aunt and Lucien, whose constant support have led me here.

And of course, thanks to Snježana, who strongly supported me during the preparation of this thesis and spent many long evenings alone during the delivery of this work; and to my little Aliénor, who was wise enough to patiently wait to the end of this thesis to discover the world.

Remerciements

Je tiens tout d'abord à exprimer ma reconnaissance à Darren Robinson, Claude-Alain Roulet et Jean-Louis Scartezzini pour m'avoir accordé leur confiance et donné la possibilité de commencer une thèse au Laboratoire d'Energie Solaire et de Physique du Bâtiment, où j'ai eu le privilège de passer de belles années de recherches fructueuses dans une équipe qui allie compétence, savoir-faire et gentillesse.

Je remercie tout particulièrement mon directeur de thèse Darren Robinson, qui m'a proposé un sujet de recherche pour lequel j'ai développé beaucoup d'intérêt. Tout au long de la préparation de cette thèse, il a apporté une supervision efficace par ses conseils avisés, en sachant clairement définir les priorités de recherche par sa vision globable du sujet. Ce travail a largement bénéficié de sa relecture soigneuse, rigoureuse et impitoyable.

J'exprime toute ma gratitude aux membres du jury, Anton Schleiss (Laboratoire de constructions hydrauliques, EPFL), Michel Bierlaire (Laboratoire transport et mobilité, EPFL), Michael A. Humphreys (Département d'architecture, Université Brookes d'Oxford) et Christoph Reinhart (Centre pour l'environnement, Université de Harvard) pour leur lecture attentive de cette thèse, leur participation à l'examen et leurs commentaires constructifs dont cette thèse a beaucoup profité.

Les données essentielles au développement des modèles de cette thèse n'auraient pas été disponibles sans le travail remarquable des collaborateurs qui m'ont précédé au laboratoire, notamment David Lindelöf, Antoine Guillemin et René Altherr à qui j'exprime ma grande reconnaissance. Je remercie également Laurent Deschamps pour sa participation à la maintenance de ce système de mesures.

J'ai également apprécié l'aide technique de Pierre Loesch et Lise Azzi pour la préparation des campagnes de mesures. Leur mise en place a été grandement facilitée par la collaboration de Christian Trachsel et Robert Sandoz (Section de l'Urbanisme, Ville de Neuchâtel), Fulvio Faralli (Service des bâtiments, Canton de Neuchâtel), Raynald Golay (EOS) et Toni Schroeter (Administration communale de Payerne). Ces campagnes de mesures n'auraient pas pu voir le jour sans l'aimable participation d'un grand nombre de volontaires, trop nombreux pour être tous nommés ici, et que je remercie pour leur aide précieuse et capitale.

Mes premiers pas dans la recherche ont grandement bénéficié des conseils de David Lindelöf qui m'a initié avec patience aux méthodes de la statistique et de l'analyse de données ainsi qu'aux outils numériques associés. Plusieurs discussions intéressantes avec Michel Bierlaire, David Daum, Antoine Guillemin, Michael A. Humphreys, Jérôme Kämpf, Ardeshir Mahdavi, Jessen Page, Claus Pröglhöf et Claude-Alain Roulet ont également aidé à développer ces travaux.

Je remercie chaleureusement les secrétaires de notre laboratoire Suzanne l'Eplattenier, Sylvette Renfer et Barbara Smith pour leur aide en tous genres et leur charmante compagnie.

Un grand merci à Lise Azzi, André Kostro et David Daum qui ont partagé mon bureau plusieurs mois avec une bonne humeur qui a illuminé mon quotidien de chercheur, même lorsque la phase de rédaction de cette thèse me rendait quelque peu taciturne. Ma gratitude va également à tous mes collègues du laboratoire qui contribuent à en faire un lieu de travail et d'étude où l'on se sent bien. J'ai une pensée particulière pour mes collègues doctorants Apiparn Borisuit, Natalia Filchakova, Mario Germano, Martin Joly, Jérôme Kämpf, Philippe Leroux, Friedrich Linhart, Diane Perez, Maria Cristina Munari Probst, Adil Rasheed, Anothai Thanachareonkit, Paola Tosolini et Urs Wilke, avec lesquels j'ai passé de bons moments au cours d'innombrables pauses cafés et joyeuses fondues.

Ce travail de recherche n'aurait pas été possible sans le soutien financier reçu tout d'abord par le Fonds National de la Recherche (54e Programme national de recherche, développement durable de l'environnement construit) et ensuite par la Commission européenne (CONCERTO II, Projet HOLISTIC), que je remercie cordialement.

Mes amis m'ont beaucoup aidé à garder un bon équilibre pendant cette période de doctorat. Je remercie notamment Olivier Anselmo, Emmanuelle Birraux, Cheryl Chang, Cynthia Dufaux, Christophe Etienne, Julien Ferdinand, Daniel Hairer, Fadi Hamdan, Brigitte Helgen, Khoa Hoàng, Gilles Humbert-Prince, Sara Jones, Aris Khan, Olivier Loutan, Gabrielle Payot, Benjamin Rappaz, Xavier Righetti, Laetitia Stroudinsky, Frank Souiller, Matthieu Sturzenegger, Christophe Tournaire, Julien Vinzio, Pascal Wehrli et Jérôme Wirz pour leur amitié et le bon temps passé ensemble. Un merci particulier à Christian Theiler qui m'a beaucoup aidé à relativiser mes soucis de doctorant au quotidien.

Quels mots pourraient suffire à exprimer ma reconnaissance à mes parents, ma chère tata et Lucien, dont les années de constant soutien m'ont conduit jusqu'ici ?

Et bien sûr, merci à Snježana, qui m'a infailliblement soutenu pendant toute la rédaction de ce travail et qui a passé beaucoup de longues soirées toute seule lors de la phase finale de l'accouchement de cette thèse, et à ma petite Aliénor qui a eu la sagesse d'attendre patiemment le point final de ce travail pour venir au monde.

Contents

1 Introduction **5**
1.1 General context of the research . . . 5
1.1.1 Energy in buildings . . . 5
1.1.2 Building and urban simulation . . . 7
1.2 Our hypothesis . . . 9
1.3 Structure of this research work . . . 9

2 The field survey **11**
2.1 The LESO-PB field survey . . . 11
2.1.1 The LESO-PB building . . . 11
2.1.2 Experimental design and measurements . . . 13
2.2 Integration of meteorological data . . . 20
2.2.1 Weather stations . . . 20
2.2.2 Calibration of outdoor temperature . . . 20
2.2.3 Calibration of wind speed and direction . . . 21
2.3 Statistical summary . . . 21
2.3.1 Descriptive statistics . . . 21
2.3.2 Representativeness of questionnaire answers . . . 22

3 Behavioural modelling **27**
3.1 Determinism and probabilistic models . . . 27
3.1.1 Some thoughts . . . 27
3.1.2 Probabilistic models in building simulation . . . 28
3.2 Generalised linear models and logistic regression . . . 29
3.2.1 Theoretical background . . . 29
3.2.2 Assessment of statistical significance . . . 32
3.2.3 Measures of goodness-of-fit . . . 33
3.2.4 Variable selection procedure . . . 34
3.2.5 Model diagnostics . . . 35
3.2.6 Ordinal logistic regression . . . 35
3.3 Random processes . . . 36
3.3.1 Bernoulli processes . . . 36
3.3.2 Markov processes . . . 36
3.4 Survival analysis . . . 37

4 Modelling actions on windows **39**
4.1 Introduction . . . 39
4.1.1 State of the art . . . 39
4.1.2 Key advances, open questions and research needs . . . 42
4.2 Results . . . 44
4.2.1 Preliminary observations . . . 44
4.2.2 Models based on single probability distributions . . . 49
4.2.3 Model based on a discrete-time Markov process . . . 52
4.2.4 Continuous-time random process . . . 58
4.2.5 Integration of individual behaviours . . . 61
4.2.6 Use of several windows . . . 67
4.3 Discussion . . . 68
4.3.1 Summary . . . 68
4.3.2 Cross-validation . . . 69
4.3.3 Treating opening angles . . . 77
4.4 Conclusion . . . 77

5 Modelling actions on shading devices **79**
5.1 Introduction . . . 79
5.1.1 State of the art . . . 79
5.1.2 Perspectives . . . 81
5.2 Patterns of actions on blinds . . . 81
5.2.1 Actions and occupancy-related effects . . . 81
5.2.2 Observed shaded fractions . . . 82
5.2.3 Variables influencing the state of blinds . . . 82
5.2.4 Stimuli for action . . . 84
5.3 Model for actions on blinds . . . 85
5.3.1 Actions on arrival on lower blinds . . . 88
5.3.2 Actions during presence and at departure on lower blinds . . . 88
5.3.3 Choice of lower blind position . . . 90
5.3.4 Model for actions on upper blinds . . . 92
5.4 Predictive accuracy . . . 92
5.4.1 Modelling variants . . . 92
5.4.2 Criteria . . . 92
5.5 Discussion . . . 96
5.6 Individuals' behaviours . . . 99
5.7 Conclusion . . . 102

6 Environmental comfort **105**
6.1 State of the art . . . 105
6.1.1 Thermal comfort . . . 106
6.1.2 Visual comfort . . . 111
6.2 Preliminary observations . . . 112
6.2.1 Statistical summary . . . 112
6.2.2 Application of the adaptive model . . . 113
6.3 Predicting thermal comfort . . . 116
6.3.1 Distribution of thermal sensation . . . 116
6.3.2 Thermal comfort probability . . . 117

6.3.3 Individual behaviours . . . 119
6.4 Comfort temperature and adaptive actions . . . 125
6.4.1 Effects of adaptive actions on thermal sensation: empirical adaptive increments . . . 125
6.4.2 Individual variations and expanded model . . . 129
6.5 Predicting visual comfort . . . 133
6.5.1 Distribution of visual sensation . . . 133
6.5.2 Visual comfort probability . . . 133
6.5.3 Individual behaviours . . . 135
6.6 Linking actions and comfort . . . 135
6.6.1 A general formulation of human adaptive actions . . . 135
6.6.2 Comfort feedback of adaptive actions . . . 138
6.7 Discussion and perspectives . . . 140

7 Integration with building simulation software 143
7.1 Implementation in building simulation . . . 143
7.1.1 Implementation of the window model . . . 143
7.1.2 Implementation of the blinds model . . . 146
7.1.3 Integration of complementary stochastic models . . . 148
7.2 Implementation in urban simulation . . . 151
7.3 Towards an agent-based modelling paradigm . . . 153
7.3.1 Simulation steps . . . 153
7.3.2 Modelling of agents' diversity . . . 155

8 Conclusion 157

A Adaptation of personal characteristics 161
A.1 Clothing . . . 161
A.1.1 State of the art . . . 161
A.1.2 Results . . . 163
A.1.3 Discussion . . . 167
A.2 Metabolic activity . . . 168
A.2.1 State of the art . . . 168
A.2.2 Results and discussion . . . 169
A.2.3 Use of drinks . . . 170

B Comfort and its interactions 173
B.1 Previous research . . . 173
B.2 Observations . . . 173
B.3 Discussion . . . 175

C Productivity of occupants 177
C.1 Introduction . . . 177
C.2 Results . . . 178
C.3 Discussion . . . 180

Chapter 1

Introduction

Diese Griechen haben sich die längste Zeit ihrer Götter bedient, gerade um sich das schlechte Gewissen vom Leibe zu halten.
These Greeks for the longest time used their gods for the very purpose of keeping the bad conscience at a distance.
Friedrich Nietzsche, On the Genealogy of Morality (1887)

1.1 General context of the research

Trends in energy demand and supply have attracted increasing concern during the last decades, to the point that the security of energy supply is now a top government priority, particularly for countries with limited fossil fuel resources. Based on projections of future energy production linked with global economical and population growth, it is today widely believed that significant structural changes must be made to achieve a more sustainable development, in societal, economical and environmental aspects. Among numerous measures to support that, the reduction of energy use through its efficient and parsimonious use is of central importance.

1.1.1 Energy in buildings

Facts and figures

The building sector is a particularly important consumer of energy. Indeed, estimations report that buildings account for around half of the total energy consumption in Switzerland and many European countries, while the specific contribution of households generally accounts for more than a quarter [1]. Their part in the global energy demand has steadily increased during the last decade [2], to the point that it has now exceeded traditionally important consumers such as transportation and industry, as can be seen in Figure 1.1. This increase is caused by several concomitant factors, including the increasing use of active cooling and electrical appliances and improved living standards.

Therefore, improving the energy performance of buildings is becoming increasingly important. However, to provide sound guidance on how to reduce buildings' energy demands requires a sound basis for predicting buildings' energy performance.

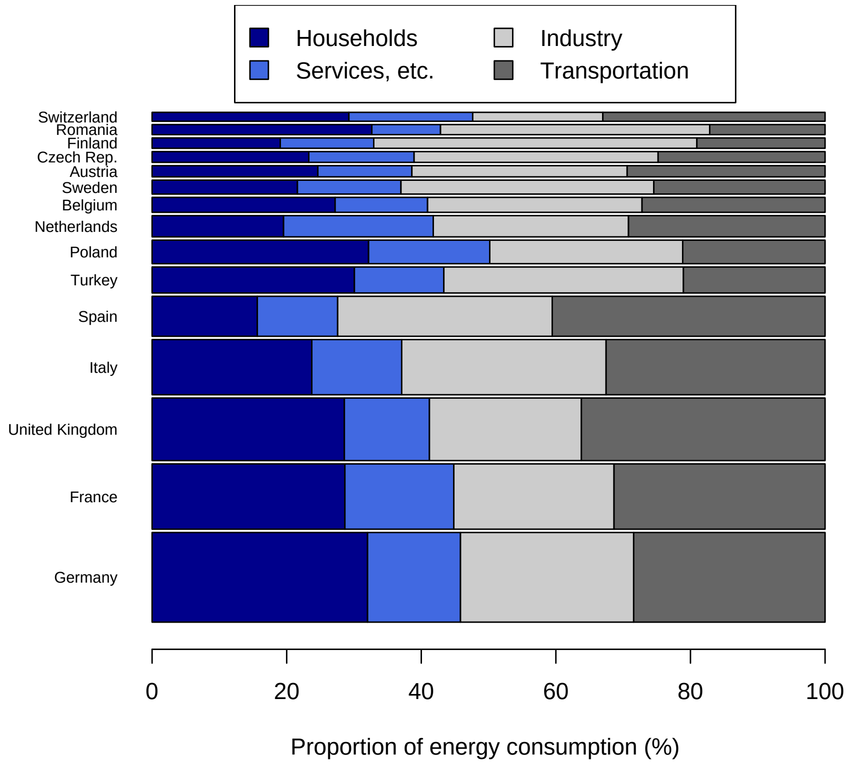


Figure 1.1: Proportion of energy consumption by sector in 2005 in the 15 most energy consuming European countries: width is proportional to total energy consumption (Source: European Union Energy and Transport in Figures [1])

Integrated building performance simulation

The large number of simultaneous objectives in building design coupled with the complexity of the energy flows of a building underlines the needs for rigorous scientific methods to support the development of low-energy buildings. In this context, the use of integrated building performance simulation programs based on accurate models is key to providing a scientific answer to the challenge of reducing energy consumption in the built environment.

However, the appropriate approach to solve these problems has to be integrated with connected domains. Buildings indeed have a footprint on the territory, their location influences the needs in transportation, they may host renewable energy technologies, buildings influence each other in daylight availability and have an impact on the urban microclimate. Therefore, considering simultaneously the interactions between all related aspects avoids unintended consequences of seemingly sensible actions and allows more reliable predictions, although the task is of huge complexity. Furthermore, changes in the scale of analysis to that of urban collectivities, allowed for by current computing capabilities, is of particular interest; so that models allowing us to consider issues such as the ecological footprint, transportation capabilities and waste management are on the horizon.

Indoor environment quality

A building is however primarily a space to live, rather than an energy consumer. Energy efficiency in buildings cannot thus be implemented by sacrificing indoor environment quality and occupant comfort – the primary purpose of a building. Unsatisfactory indoor environment quality may provoke unintended consequences, such as the expansion of uncontrolled energy consuming air-conditioning systems. Furthermore, low indoor environment quality induces a large range of direct problems, such as sick building syndrom (SBS), lower general wellbeing and satisfaction and lower productivity at work [3, 4, 5, 6].

On the other hand, there is an increasing demand from the public for high quality living space. The challenge is to ensure high indoor environment quality with minimal energy consumption. Strategies to achieve this compromise – where the use of daylight and the possibilities to act on the environment play a central role – exist and must be integrated at the design stage. As such our simulation tools need to be able to predict not only energy flows but also the quality of the indoor environment.

1.1.2 Building and urban simulation

In this section, we review the developments already performed in the domain of integrated building performance simulation and point out the central research needs to increase the accuracy of simulation tools.

Current features of dynamic building simulation tools

The first generation of dynamic simulation programs was developed during the 1970s and early 1980s [7, 8]. These were essentially command-line interfaces to routines to calculate the dynamic thermal energy exchanges within a building and between this and the outside environment. Subsequent work concentrated on improving the usability of these routines and extending the scope of the core capabilities, for example to incorporate coupled plant [9] and mass flow [10] modelling.

With improved functionality and amidst growing demand for their use by the more pioneering design consultants, attention shifted to proving the validity of their core thermal energy exchange models [11, 12, 13]. By the mid 1990s, with results from these validation studies taken on board and with improved usability, attention then focused upon the addition of further modelling functionality. This included the addition of 3D conduction modelling [14], links with ray tracing programs for improved lighting modelling [15], electrical power flow modelling [16], embedded computational fluid dynamics (CFD) [17] and sensitivity analysis [18, 19].

The results are programs such as ESP-r [20] with a transient finite difference heat flow solver at the core, supporting simultaneous solutions of plant, fluid, electrical power and CFD equation sets.

The central role of buildings occupants

The deterministic features of building simulation programs are now considered relatively mature. But their ability to emulate reality is undermined by a poor representation of non-deterministic variables, particularly relating to occupants presence and their interactions with environmental controls.

Indeed discrepancies between real and assumed deterministic behaviour are such that predictions of like buildings may, in Baker's estimation, vary by a factor of two [21], an estimation confirmed by field observations[1]. Considering ventilation, measurements conducted in 25 Danish buildings showed that on average the increase in the mean airflow rate due to the influence of occupancy is more than 100% [23]. Iwashita and Akasaka [24] later measured that 87% of the total air change rate is caused by the behaviour of the occupants. More recently, Bahaj and James [25] observed that the electricity consumption in 9 identical low-energy social housing units varied by as much as 600% in some periods of the year[2].

According to Hoes and Hensen [26], the relative influence of the users' behaviour increases in passive buildings, which are expected to become more common due to the demand for sustainable buildings. They point out that for some buildings detailed behavioural modelling is necessary to design buildings that are robust to the influence of user behaviour.

To better understand the nature of the problem we list here a range of types of interaction available to occupants and their respective influences on buildings' performance:

- **Presence of occupants.** Occupancy is the essential condition governing any further action, as the occurrence of an interaction needs an occupant to perform it. Moreover, occupants are a direct source of metabolic heat gains, humidity, CO_2 and pollutants.
- **Actions on windows.** Window openings and their associated air flows have an important impact on indoor hygro-thermal conditions and indoor air quality (eg. concentration of pollutants) particularly in naturally-ventilated buildings.
- **Actions on doors.** Open doors favour air flows inside buildings and thus the transport of heat and pollutant. They may also amplify the air flows from open windows through cross-ventilation.
- **Actions on shading devices.** The position of shading devices determines the solar heat gains, with a corresponding impact on the evolution of indoor temperature. Furthermore, they directly determine daylight availability and the transmitted luminous flux through the windows of a building.
- **Use of artificial lighting.** The primary interest of modelling lighting use is in the prediction of electrical energy consumption (and associated heat gains). Furthermore, they may be interrelated with actions on shading devices.
- **Use of heating, ventilation and air-conditioning (HVAC) systems.** Actions on these devices directly impact energy consumption and indoor conditions.
- **Use of water and electrical appliances.** Together with the use of lighting, the prediction of appliance use enables estimation of the total electricity consumption (as well as useful statistics like base and peak load) and the corresponding space heat gains, with an extension to water use prediction if associated appliances are in use.
- **The production of waste.** Occupancy and appliances use may be combined to predict the production of waste water and solid waste, from which energy may be derived.

[1]For example Seligman et al. [22] investigated energy consumption in 28 identical town houses and found that the largest variation in energy consumption was two to one.

[2]In the same study, an indicator such as daily consumption feedback provided to occupants was found to reduce electricity usage by 10%, while the device signaling homeowners when they could cool their houses without air conditioning by opening their windows led to a reduction in consumption of 15.7%.

Furthermore, other behavioural patterns have a possible influence on the occurence of these actions, so that they are also of indirect relevance. In particular it may be useful to consider modelling:

- **Actions on fans.** Occupants may use fans to maintain acceptable thermal comfort during a heat wave. They have furthermore a marginal impact on electricity use.
- **Personal characteristics.** Clothing level, metabolic activity, and cold or hot drinks consumption have an impact on thermal comfort and thereby may also influence subsequent actions of interest.

Although the significant impact of occupants' actions on building thermal conditions and energy flows has longsince been established, algorithms to simulate them are either inexistant or based on limited assumptions. ESP-r is among the rare software currently including some behavioural control models, for example for the dynamic control of lighting, shading devices and windows. However, we will show in Chapters 4 and 5 that these latter models are not fully satisfactory. Their calibration basis is also very limited. In the current state of research, we suggest that rigorously derived and validated models have only been developed to simulate occupancy, based on the research of Wang [27], Page [28, 29] and Richardson [30]. We will defend this assertion in later chapters.

With respect to integrated simulation at the urban scale, the aggregated results of stochastic models of behaviour are of particular interest for sizing of energy supply and storage infrastructure.

1.2 Our hypothesis

Current approaches in building performance simulation are undermined by important flaws with respect to deficiencies in the modelling of buildings occupants' behaviour. But occupants' behaviour is complicated by the fact that their perception of environmental comfort influences their actions. This issue has thus far been ignored in building simulation. Considering these insufficiencies, we propose the following hypothesis (also shown in Figure 1.2) to guide our developments:

> **The dynamic thermal performance of buildings is intrinsically linked to the behaviour of its occupants, which is itself directly influenced by perceived environmental comfort. Furthermore, actions taken by occupants to restore their environmental comfort have a quantifiable physical, physiological and possibly psychological feedback on this latter, which is of central importance for satisfaction with the indoor environment.**

1.3 Structure of this research work

This work begins with a detailed description of our measurement campaign and its experimental design (Chapter 2) which was developed to address the above hypothesis. We then present the key mathematical principles of relevance to behavioural modelling (Chapter 3).

We go on to propose modelling approaches to describe relevant behavioural patterns of buildings' occupants. In Chapter 4 we develop an appropriate stochastic model for the

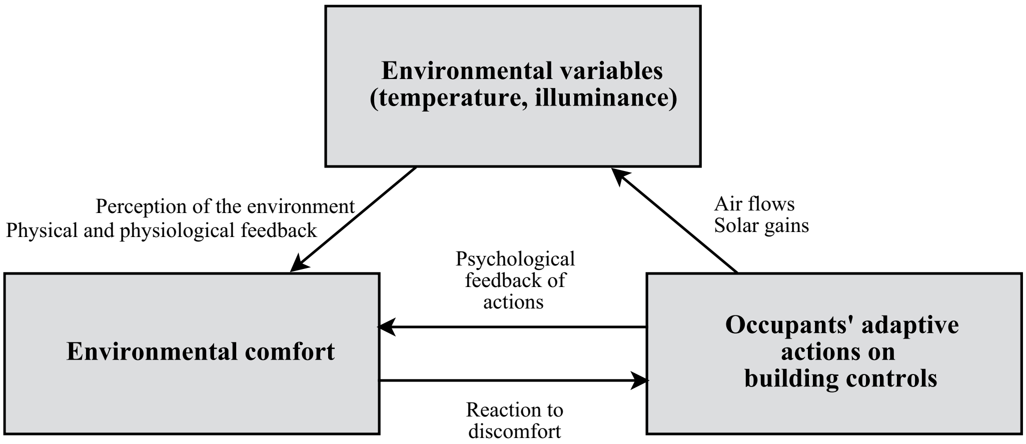


Figure 1.2: General scheme of the relationships proposed in our hypothesis

predictions of occupants' actions on windows, followed in Chapter 5 by a similar procedure to predict actions on shading devices.

As a next step, we question in Chapter 6 the appropriateness of the currently accepted model for thermal comfort and go on to propose new extensions based on our hypothesis and show the corresponding improved accuracy and conceptual rigour. This results in an innovative way to formulate the stochastic models for actions as a direct consequence of environmental discomfort, where the notion of action inertia expresses this link.

Finally, we present in detail an appropriate approach to integrate these findings and describe in detail how to implement the results in building and urban simulation tools. We also discuss the value of an agent-based approach in this context (Chapter 7).

This work is completed by three appendices where we present research performed in connected domains: adaptation of personal characteristics such as clothing and metabolic activity (Appendix A); a discussion of the interaction between thermal, olfactory and visual comfort based on experimental evidence (Appendix B); and an analysis of factors influencing productivity in office environments (Appendix C).

Chapter 2

The field survey

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.
John W. Tukey (1915-2000), Sunset salvo [31]

This chapter presents the experimental design that provided the basis for the development of our models. We present first the general characteristics of the surveyed building and go on to describe the measurement devices, surveyed periods and details of the experimental design. We conclude with a statistical summary of the measured data including initial exploratory data analysis and relevant preliminary observations for further model development.

2.1 The LESO-PB field survey

2.1.1 The LESO-PB building

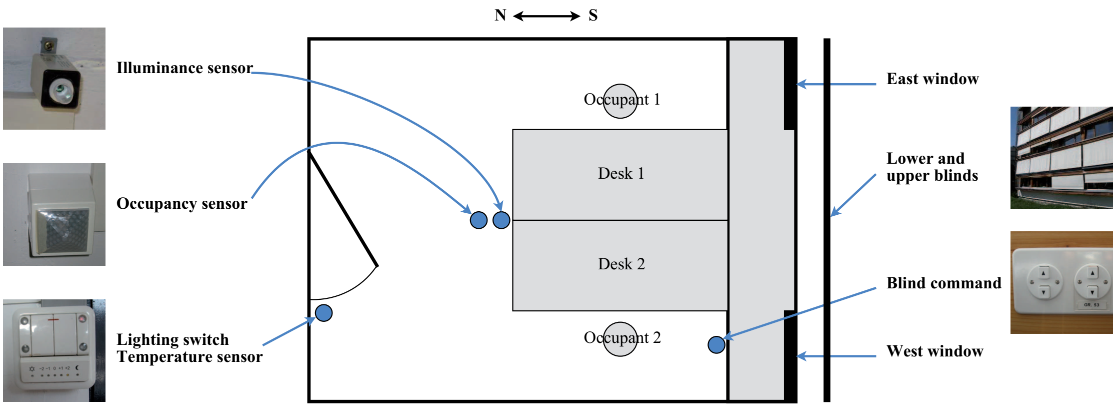
The Solar Energy and Building Physics Laboratory (LESO-PB) experimental building (Figure 2.1(a)), located in the suburb of Lausanne, Switzerland (46°31'17"N, 6°34'02"E, alt. 396 m.) was built in 1982 and renovated in 1999. It hosts on three floors fourteen south-facing offices, a workshop, a conference room, a computer room and a small library. The building has no mechanical ventilation system. All south facing offices have an area of 15.7 m^2 and a height of 2.8 m; they are equipped with anidolic systems that improve the distribution of daylight. A typical office is shown in Figure 2.1(b).

In every south-facing office, occupants have the possibility to tilt or open up to any angle each of the two windows (height 90 cm, width 70 cm). It is safe to leave windows open (eg. for night ventilation) during periods of absence, except on the ground floor. Occupants also have the possibility to control two external blinds (width 350 cm): a lower blind potentially covering the totality of the vision[1] window (height 100 to 185 cm) and an upper blind covering the anidolic system (height 210 to 270 cm). These blinds are controlled by switches (one to start and one to stop lowering/raising) which allow occupants to shade their windows to any desired fraction. Occupants may also close and tilt internal vertical slat blinds at the upper window to reduce glare whilst benefitting from direct solar gain during the heating season. Electrical lighting may be dimmed to any desired level with

[1]The lower part accounts for the main part of outdoor visibility, see Figure 2.1(b).

(a) View of the south façade of the LESO building

(b) Typical cellular office of the LESO building



(c) Scheme of the cellular offices showing sensors and controls

Figure 2.1: Features of the LESO building

a switch (Figure 2.3(d)) located next to each office's door. Finally, several offices are equipped with ceiling fans.

Occupants of this building include senior researchers, research assistants, technical staff and secretaries. They mainly carry out office related work and all use a computer. A majority of occupants occupy long-term positions in the laboratory. During the surveyed period, between five and eight offices have been occupied by two persons (Table 2.5), which can both individually access their own window, while between six and nine offices have accommodated single occupants who are able to act on the two windows.

A detailed description of the building with an exhaustive analysis of the building's energy flows, is provided by Altherr and Gay [32].

Throughout this work we will refer to offices by three numbers (001 to 004, 101 to 106 and 201 to 204), where the first number indicates the floor. Individual occupants with known presence periods are attributed a number from 1 to 43.

2.1.2 Experimental design and measurements

Permanent measurements. Since December 2001, all 14 south-facing cellular offices of this building have been progressively equipped with sensors whose real-time measurements are archived by a centralised EIB data acquisition system. Figure 2.1(c) describes the position of these sensors and Table 2.1 lists the variables measured along with their respective periods of availability (with the exception of a few interruptions caused by maintenance and technical reasons). Figure 2.2 shows as an example the measurements performed during the period 6$^{\text{th}}$ to 12$^{\text{th}}$ October 2008.

The following variables have been continuously[2] measured up until the final data collection on 8$^{\text{th}}$ September 2009,:

- Occupancy was measured by presence infrared detectors (Figure 2.3(a)). The number of occupants cannot be recorded by the sensors.

- Window openings and closings were detected by micro switches on each of the two windows (Figure 2.3(b)). These devices do not record the opening angle and do not distinguish fully open from tilted openings.

- The position of lower and upper blinds (B_L and B_U) was recorded after every observed movement through the measured extension of the cables supporting them. Occupants act on blinds through commands shown in Figure 2.3(c).

- The status of electrical lighting (L) was measured directly from the switches (Figure 2.3(d)).

- Local indoor (θ_{in}) and roof outdoor (θ_{out}) temperature was measured by Pt-100 resistance thermometers (Figure 2.3(e)). In the offices, these devices were enclosed in the light switches (Figure 2.3(d)), whilst the roof sensor is housed within a Stevensen screen.

- Indoor horizontal workplane illuminance (E_{in}) was measured by Siemens brightness sensors GE 252 (Figure 2.3(f)) protected from the window's luminance.

[2] The acquisition system records the changes observed by measurement devices in real time. Temperature and illuminance are recorded along with a time stamp once the variation exceeds 0.06°C or 30 lx.

Measurement	Symbol	Unit	Start	End	Dur.	Freq.
Direct measurements						
Indoor temperature	θ_{in}	[°C]	18.12.2001	08.09.2009	2814 d.	Cont.
Outdoor temperature	θ_{out}	[°C]	16.03.2005	08.09.2009	1630 d.	Cont.
Indoor illuminance	E_{in}	[lx]	24.09.2003	08.09.2009	2169 d.	Cont.
Outdoor global horiz. illuminance	$E_{\mathrm{gl,hor}}$	[lx]	11.07.2003	08.09.2009	2245 d.	Cont.
Outdoor global horiz. irradiance	$I_{\mathrm{gl,hor}}$	[W/m^2]	11.07.2003	08.09.2009	2245 d.	Cont.
Outdoor diffuse horiz. irradiance	$I_{\mathrm{diff,hor}}$	[W/m^2]	19.02.2005	08.09.2009	2245 d.	Cont.
Occupancy status	O	Boolean	18.12.2001	08.09.2009	2814 d.	Cont.
Windows status (east and west)	W_E, W_W	Boolean	18.12.2001	08.09.2009	2814 d.	Cont.
Lighting status	L	[%]	18.12.2001	08.09.2009	2169 d.	Cont.
Lower blind shaded fraction	B_L	[%]	01.01.2003	08.09.2009	2169 d.	Cont.
Upper blind shaded fraction	B_U	[%]	01.01.2003	08.09.2009	2169 d.	Cont.
Deduced						
Sun altitude	ζ	[°]	18.12.2001	08.09.2009	2814 d.	Cont.
Sun azimuth	α	[°]	18.12.2001	08.09.2009	2814 d.	Cont.
Added devices						
Local temperature	θ_{loc}	[°C]	23.06.2008	06.09.2009	440 d.	10 min.
Indoor humidity	ϕ_{in}	[%]	11.09.2007	06.09.2009	726 d.	10 min.
Questionnaire						
Ceiling fan status	F	Boolean	13.06.2006	08.09.2009		ca. 2 h.
Clothing level	I_{cl}	[m^{2}K/W]	13.06.2006	08.09.2009		ca. 2 h.
Specific metabolic activity	M	[W/m^2]	13.06.2006	08.09.2009		ca. 2 h.
Thermal sensation	S_{th}	7 levels	13.06.2006	08.09.2009		ca. 2 h.
Visual sensation	S_{vis}	7 levels	13.02.2008	08.09.2009		ca. 2 h.
Olfactory sensation	S_{olf}	7 levels	13.02.2008	08.09.2009		ca. 2 h.
Hot and cold drinks	D_H, D_C	Boolean	13.06.2006	08.09.2009		ca. 2 h.
Perceived productivity	P	[%]	13.02.2008	08.09.2009		1 d.

Table 2.1: List of the parameters measured during the LESO-PB survey with their measurement periods, total durations and intervals

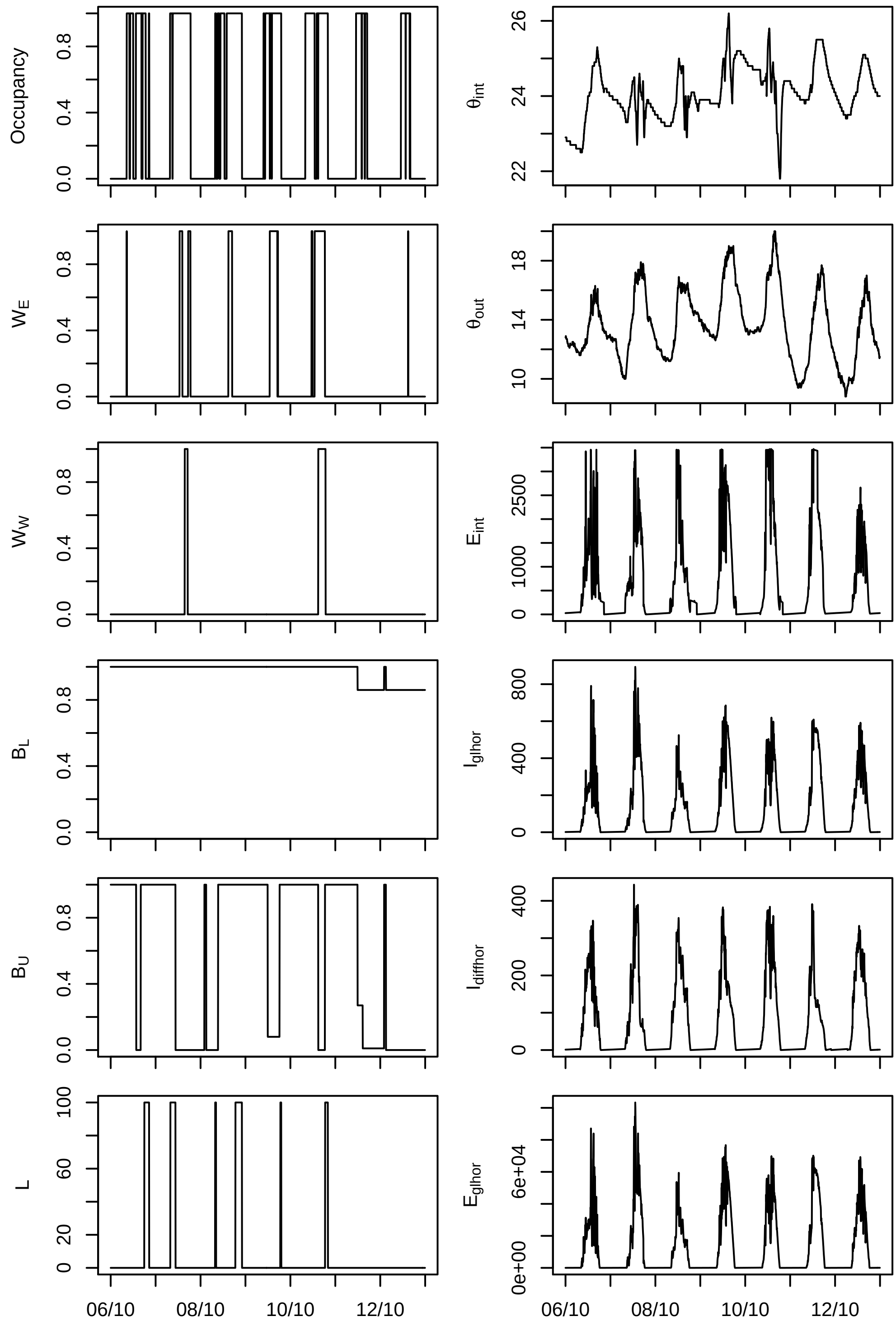


Figure 2.2: Measurements performed in the office 002 during the week from Monday 6 to Sunday 12 October 2008

- Outdoor global horizontal illuminance ($E_{\mathrm{gl,hor}}$), outdoor global ($I_{\mathrm{gl,hor}}$) and diffuse ($I_{\mathrm{diff,hor}}$) horizontal irradiance were measured by a Delta-T BF3 sunshine sensor (Figure 2.3(g)) installed on the roof.

Finally, we computed sun elevation (ζ) and azimuth (α) using the Astronomical Almanac algorithm[3].

Local indoor temperature and relative humidity. We placed in some offices temperature Tinytag TG-0050 sensors (Fig. 2.3(h)) in the immediate vicinity of occupants to assess a potential difference with measurements from the central system. Furthermore, we measured indoor humidity (ϕ_{in}) through Tinytag TK-4014 sensors (Fig. 2.3(i)) in a few offices for distinct periods of time.

Electronic questionnaire. An electronic questionnaire (Figure 2.4(a)), developed using Borland Delphi programming environment, was activated on the computers of all occupants on a rotation basis. When possible, each occupant was surveyed for at least three separate periods of three months in winter, in summer and during an inter-seasonal period. The questionnaire typically appeared four times a day, twice in the morning and in the afternoon, at intervals of between 2 and 3 hours (defined in agreement with each occupant). The program was activated by the scheduled tasks manager of the operating system.

At every prompt, occupants were asked to provide the following information:

- **Current clothing level.** The surveyed occupants could choose from amongst eight possibilities proposed in a rolling list (Table 2.2, top), from which typical clothing insulation values could be deduced from the ISO 7730 standard [35].

- **Activity level during the preceding 15 minutes.** Six possibilities (Table 2.2, bottom) were offered, also based on the ISO 7730 standard [35]. Respondents could also tick a "no change" box, following from the first prompt of the day.

- **Thermal sensation.** An approximate French translation of the standard seven point ASHRAE scale (itself used in the ISO 7730 standard) was proposed (Table 2.3).

- **Visual sensation.** To our knowledge the literature does not provide a standard scale. We proposed seven choices similar to the ASHRAE scale for thermal sensation, ranging from "Very dark" to "Very bright" and centered on "Comfortable" (Table 2.3). This scale was chosen based on the desire to propose denominations where the degree of associated discomfort is similar to the choices proposed with the thermal sensation scale. The same adverbs are thus used to characterise mild deviation from neutrality (-1 and $+1$) and strong discomfort (-3 and $+3$).

- **Olfactory sensation.** A seven-point scale was used, proposing values from "Unacceptable" to "Excellent". While thermal and visual discomfort may be caused by both lack or excess of heat and light, olfactory discomfort occurs in a unique way; which imposes the use of a unidirectional scale. The denomination of the central category is here designed to evoke an average air quality while other choices explicitly refer to notions of poorness and goodness of current conditions.

[3] The translation in R language provided by Lindelöf [33] was used for the calculations. Details of the implementation of this algorithm are presented by Seidelmann [34].

(a) Infrared detector for occupancy

(b) Micro switch recording window openings

(c) Switches for actions on upper and lower blinds

(d) Switch for electrical lights

(e) Pt-100 resistance thermometer

(f) Illuminance sensor (Siemens GE 252)

(g) Sunshine sensor for outdoor irradiance (Delta-T BF3)

(h) Portable temperature sensor (Tinytag TG-0050)

(i) Portable humidity sensor (Tinytag TK-4014)

Figure 2.3: Measurement devices used in the field survey

Clothing ensemble		Insulation [m^2K/W]	[clo]
Jacket, shirt with long sleeves, trousers/dress, tie, shoes	Veston, chemise manches longues, pantalons/robe, cravate, chaussures	0.147	0.95
Jacket, open neck shirt, trousers/dress, shoes	Veston, chemise à col ouvert, pantalons/robe, chaussures	0.140	0.90
Shirt with long sleeves, trousers/dress, tie, shoes	Chemise manches longues, pantalons ou robe, cravate, chaussures	0.124	0.80
Shirt with long sleeves, trousers/dress, shoes	Chemise manches longues, pantalons/robe, chaussures	0.116	0.75
Sweater, shirt, trousers, shoes	Pullover, chemise, pantalons, chaussures	0.109	0.70
Shirt with short sleeves, trousers, shoes	Chemise manches courtes, pantalons/robe, chaussures	0.093	0.60
Shirt with short sleeves, trousers, sandals	Chemise manches courtes, pantalons/robe, sandales	0.080	0.50
Shirt with short sleeves, short/skirt, shoes	Chemise manches courtes, shorts/jupe, chaussures	0.062	0.40
Shirt with short sleeves, short/skirt, sandals	Chemise manches courtes, shorts/jupe, sandales	0.047	0.30

Activity		Metabolic rates [W/m^2]	[met]
Seated, relaxed	Assis, inactif	58	1.0
Sedentary activity	Activité sédentaire	70	1.2
Standing, light activity	Activité légère, debout	93	1.6
Standing, medium activity	Travail debout	116	2.0
Walking	Marche	140	2.4
Cycling, running	Cyclisme, course	232	4.0

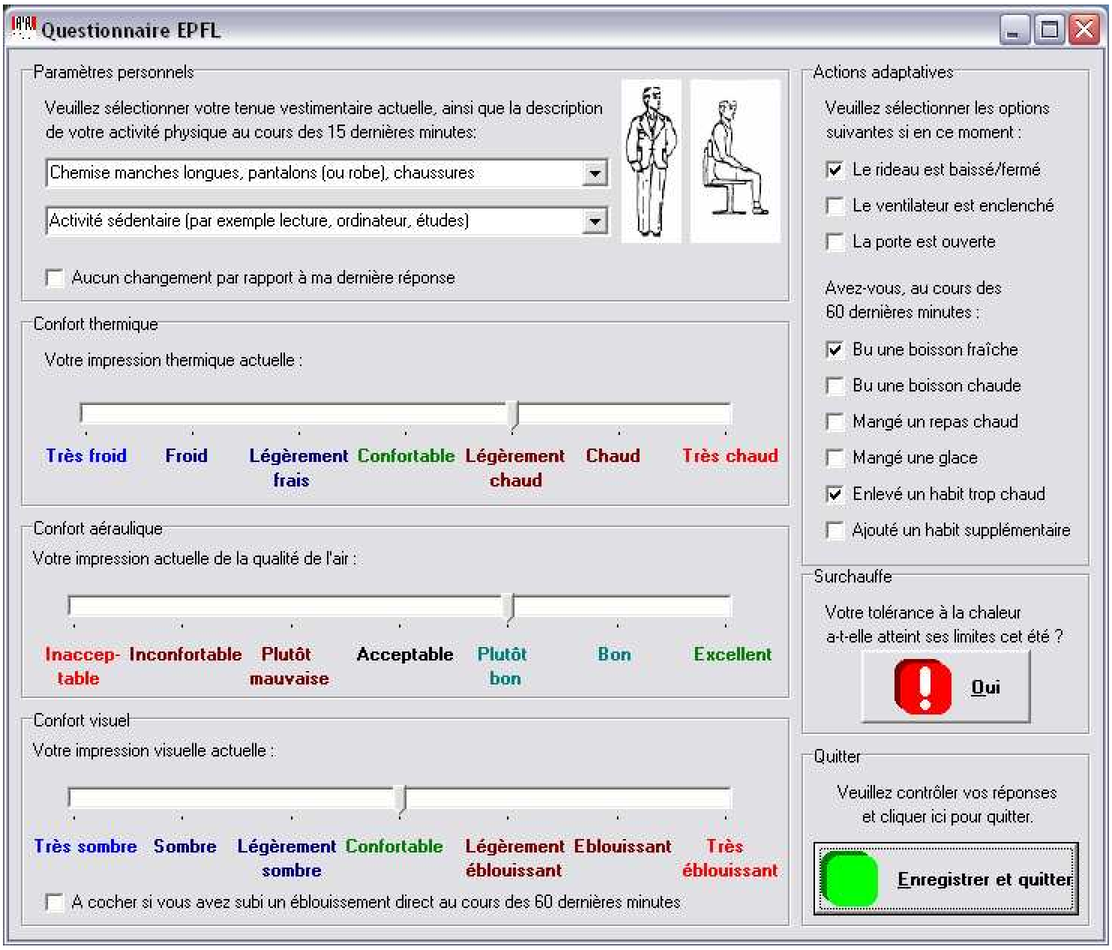
Table 2.2: Choices available in the electronic questionnaire for clothing and activity, with corresponding values from ISO-7730 Standard

- **Glare.** The occurrence of direct glare during the preceding 60 minutes.

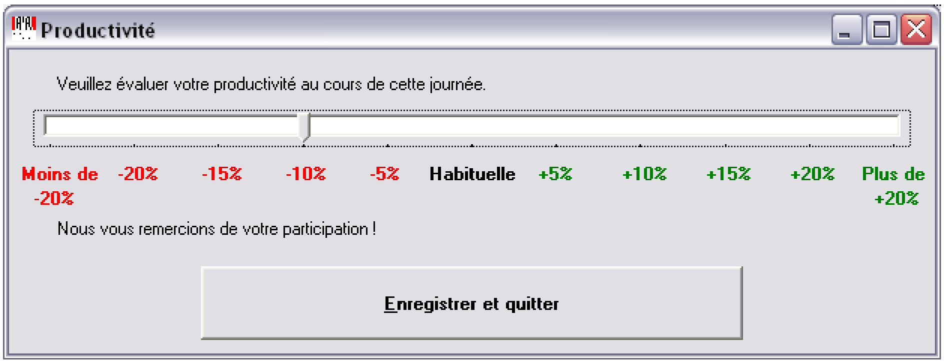
- **Other controls.** Current state of controls that were not recorded by the central system (door, ceiling fan, curtains).

- **Other activities during the preceding hour.** This includes the intake of hot and cold drinks, meals, and additions or removals of clothing items.

Furthermore, occupants were asked to estimate their productivity during the present day. This question appeared as a separate window once a day after 3pm. Available choices ranged between −20% and +20% of their usual productivity, at 5% increments (Figure 2.4(b)).



(a) Main survey window



(b) Productivity survey window

Figure 2.4: Electronic questionnaire

Scale	Thermal sensation		Visual sensation		Olfactory sensation	
-3	Cold	Très froid	Very dark	Très sombre	Unacceptable	Inacceptable
-2	Cool	Froid	Dark	Sombre	Uncomfortable	Inconfortable
-1	Slightly cool	Léger. frais	Slightly dark	Légèr. sombre	Rather poor	Plutôt mauvaise
0	Comfortable	Confortable	Comfortable	Confortable	Acceptable	Acceptable
+1	Slightly warm	Légèr. chaud	Slightly bright	Légèr. éblouissant	Rather good	Plutôt bon
+2	Warm	Chaud	Bright	Eblouissant	Good	Bon
+3	Hot	Très chaud	Very bright	Très éblouissant	Excellent	Excellent

Table 2.3: Thermal, visual and olfactory sensation levels proposed in the electronic questionnaire

2.2 Integration of meteorological data

In order to consider comprehensively the influence of outdoor parameters, it is desirable to use local meteorological data. Unfortunately however local outdoor temperature data are missing for the first three years of measurements (Table 2.1); furthermore no other outdoor climate variables were measured outside the LESO building.

In this section, we will therefore describe the characteristics of two neighbouring meteorological stations and explain the method by which we integrated their measurements into the LESO building dataset for the purposes of our study.

2.2.1 Weather stations

The first weather station which is within close proximity of the LESO building records every 10 minutes measurements of dry bulb air temperature (θ_{out}), mean wind speed (v_{wind}) and direction (α_{wind}), relative humidity (ϕ_{out}), rainfall (D_{prec}) and reduced atmospherical pressure ($p_{\text{atm,red}}$). This station is located 7.7 km away in the town Pully (46°30'43"N, 6°40'03"E, alt. 461 m.), and is part of the Meteosuisse ANETZ network, which records comprehensive measurements throughout Switzerland. The second weather station which is located 10.4 km away in Saint-Prex (46°29'01"N, 6°26'34"E, alt. 425 m.) is part of the secondary network ENET of Meteosuisse. At this station, only wind speed and direction are measured every 10 minutes.

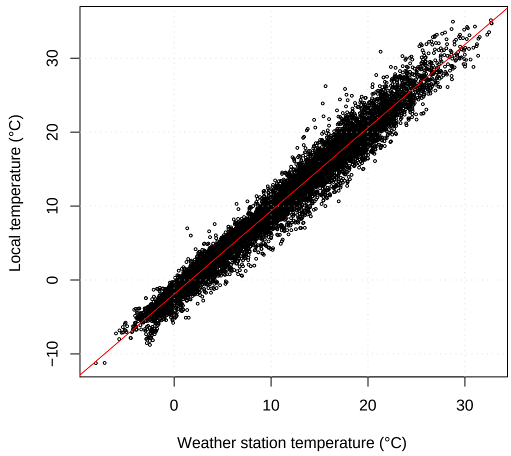
A statistical summary of the above variables is presented in Table 2.4.

2.2.2 Calibration of outdoor temperature

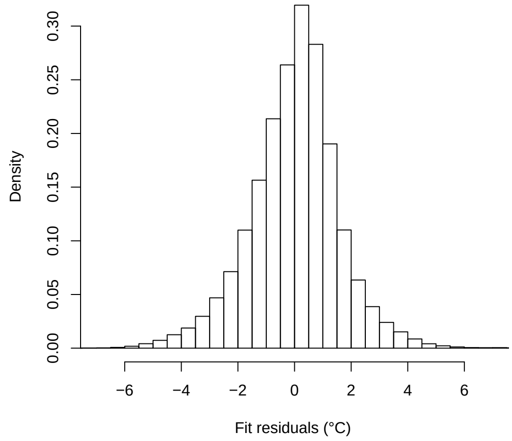
In order to extend our outdoor climate data to the first three years of measurements, we deduce a relationship between local outdoor temperature measured on the roof of the laboratory $\theta_{\text{out,loc}}$ and at the weather station $\theta_{\text{out,ms}}$ using linear regression analysis.

Using data recorded between 16 March 2005 and 9 September 2009, we notice that the median value of $(\theta_{\text{out,ms}} - \theta_{\text{out,loc}})$ is 0.56°C, with quartiles -0.63°C and 1.57°C, which indicates a systematic deviation. Performing a linear regression between these variables, we obtain $\theta_{\text{out,loc}} = a + b\theta_{\text{out,ms}}$, where $a = -1.8762 \pm 0.0027$, $b = 1.12329 \pm 0.00019$, with little dispersion ($R^2 = 0.961$), see Figure 2.5(a).

A detailed examination of the fit residuals $\varepsilon = \theta_{\text{out,loc}} - \theta_{\text{out,fitted}}$ shows that 50% of them lie within the interval $[-0.95°\text{C}, 0.96°\text{C}]$, and 95% in $[-3.72°\text{C}, 3.52°\text{C}]$. A histogram of the residuals is presented in Figure 2.5(b). We conclude thus that the discrepancy between local and distant observations has an acceptable amplitude, and we will use the adjusted weather data to extend our database.



(a) Outdoor temperature outside LESO-PB building versus weather station, with linear regression line



(b) Histogram of fit residuals

Figure 2.5: Calibration of local outdoor temperature

2.2.3 Calibration of wind speed and direction

The measurements of wind speed and direction present the additional problem of the highly local nature of observations, which undermines the relevance of more distant observations. For instance, we only obtain a correlation of 0.645 between wind speeds measured at the weather stations.

However, a strong wind measured at the meteorological station generally results in at least in some wind at the LESO building, and in this case the mean wind direction will be roughly similar. It may therefore be reasonable to use a coarse representation of wind speed and direction, by considering four levels of wind intensity defined by the observed quartiles of wind speed at the weather station given in Table 2.4: very low (< 1.5 m/s), low (1.5-2.5 m/s), moderate (2.5-4.8 m/s) and high (> 4.8 m/s). We similarly use four levels for wind direction, defined by 90° sectors centered along the cardinal points. These choices allow us to assess the existence of wind effects on occupants' behaviour (but not to quantitatively estimate its influence).

2.3 Statistical summary

2.3.1 Descriptive statistics

We present in Table 2.4 a statistical summary of the measured climatic data during occupants' presence. We also show in Figure 2.6 bivariate plots between all considered physical variables, together with their correlations and histograms for the distributions of all the variables. As expected, we observe clear positive correlations between θ_{out} and θ_{in}, as well as between θ_{out} and ϕ_{out}. It is also encouraging to note that high values of v_{wind} mostly occur for two particular ranges of α_{wind} – a characteristic phenomenon of the region of Lausanne. Finally, outdoor illuminance and direct and diffuse irradiance are also strongly correlated again as expected. However, when there are such significant correlations be-

Variable	**Min.**	**q(25%)**	**Median**	**Mean**	**q(75%)**	**Max.**
Weather station						
θ_{out} [°C]	-9.7	5.7	12.2	12.30	18.5	37.1
ϕ_{out} [%]	10.5	57.3	68.8	67.71	79.5	100.0
v_{wind} [m/s]	0.0	1.5	2.5	3.336	4.8	16.8
α_{wind} [°]	0	46	169	156.2	236	360
D_{prec} [mm/h]	0	0	0	0.1101	0	96.6
$p_{\text{atm,red}}$ [hPa]	982	1013	1018	1018	1023	1042
Roof of LESO-PB building						
θ_{out} [°C]	-10.6	4.9	11.6	11.86	18.6	36.0
$E_{\text{gl,hor}}$ [lx]	0	628	18566	29667	50708	155038
$I_{\text{gl,hor}}$ [W/m^2]	0	9.79	146.91	262.01	436.38	1270.19
$I_{\text{diff,hor}}$ [W/m^2]	0	2.44	72.14	134.07	160.72	1276.26
Aggregated measurements						
$\theta_{\text{out,dm}}$ [°C]	-7.48	4.92	10.95	11.10	17.12	28.92
$\theta_{\text{out,rm}}$ [°C]	-5.43	4.98	10.86	11.04	17.12	28.38
$\theta_{\text{out,mm}}$ [°C]	-1.45	4.69	10.56	11.07	17.65	24.68

Table 2.4: Descriptive statistics of measured parameters during the occupied intervals of the surveyed period

tween variables caution should be exercised when using them jointly as driving variables in a model, as collinearity may cause problems in the regression process (Section 3.2.4).

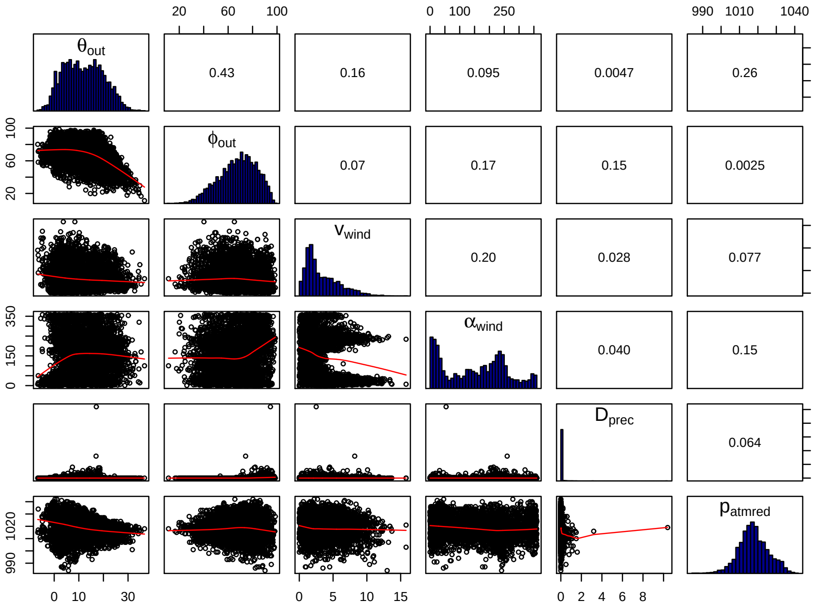
Figure 2.7 shows the distributions of indoor temperature and illuminance for all known combinations of occupants. Although all offices are identical, some modest variations are noticeable between them – a fact to be discussed later with respect to observations on environmental comfort and adaptive actions. Finally, Table 2.5 shows significant differences in the use of controls (windows, blinds and lights) between occupants, suggesting that there is considerable diversity in observed behaviours.

2.3.2 Representativeness of questionnaire answers

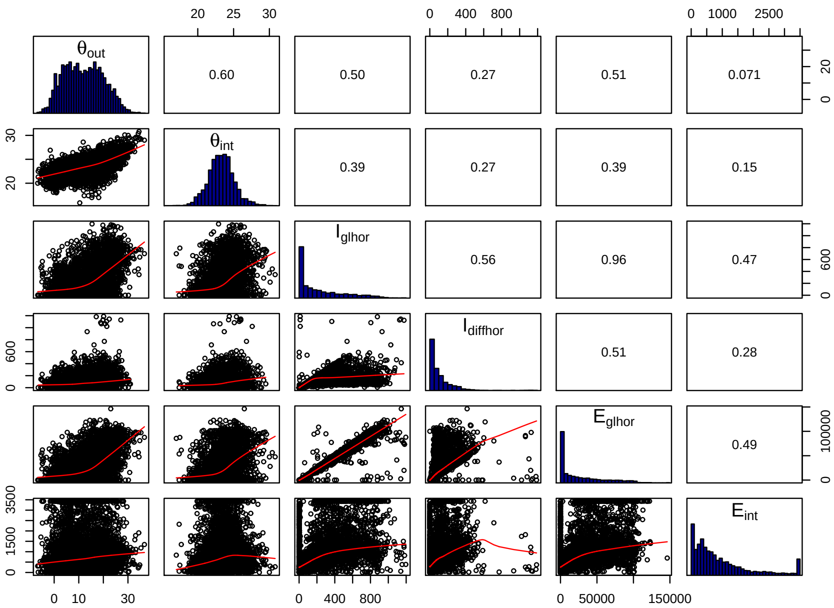
The questionnaire was completed by 28 occupants (Table 2.6), who each provided between 37 and 661 answers, totaling 6851 entries (with an average of 245 entries per person). Care was taken when administering the electronic questionnaire (Section 2.1.2) to ensure a balance in periods throughout the year. In order to confirm that the answers to the questionnaire form a representative subset of usual laboratory conditions, we check for the agreement between distributions of key variables for overall conditions and for the periods when the questionnaire was completed. If there is a good agreement then the statistical summaries from the questionnaire data may be considered to describe well the general conditions within the building.

The distributions of the day of the year, indoor temperature and indoor illuminance are shown in Figure 2.8 for the whole occupied period and for the periods corresponding to electronic questionnaire prompts. From Figure 2.8(a) we observe that inter-seasonal periods are underrepresentated, so there is some lack of generality according to seasonal patterns. However, Figures 2.8(b)-2.8(c) suggest that distributions of indoor temperature and indoor illuminance are very closely matched.

Finally, eight occupants could not be surveyed for a representative subset of normal



(a) Data from the weather station

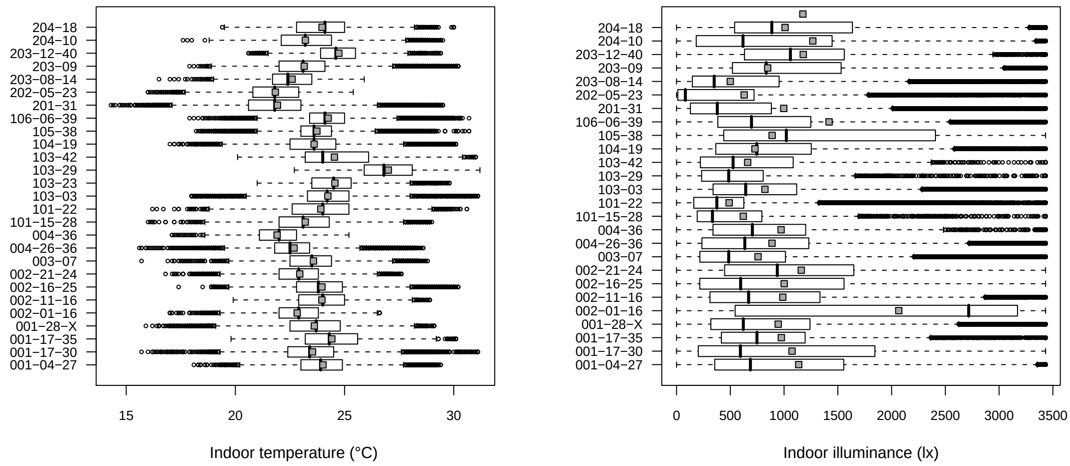


(b) Data recorded at the LESO-PB building

Figure 2.6: Bivariate plots between all measured physical variables, with local polynomial regression, correlations and histograms

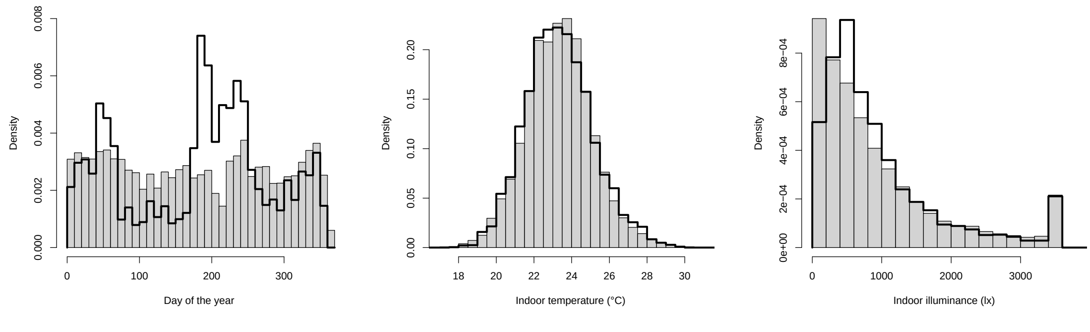
Ref.	Survey Duration (days)	Prop. of windows open East	Prop. of windows open West	Prop. of windows open Any	Lower blinds Mean frac.	Lower blinds Fully Raised	Lower blinds Fully Lowered	Upper blinds Mean frac.	Upper blinds Fully Raised	Upper blinds Fully Lowered	Prop. lights on
001-04-27	730	0.092	0.146	0.210	0.363	0.157	0.559	0.346	0.149	0.536	0.285
001-28-X	699	0.110	0.263	0.343	0.736	0.631	0.183	0.691	0.482	0.192	0.145
001-17-30	931	0.267	0.283	0.459	0.861	0.573	0.051	0.648	0.536	0.152	0.186
001-17-35	206	0.383	0.085	0.460	0.717	0.507	0.110	0.459	0.283	0.225	0.020
002-21-24	275	0.276	0.324	0.400	0.926	0.732	0.005	0.508	0.372	0.220	0.120
002-16-25	334	0.285	0.256	0.470	0.920	0.852	0.056	0.573	0.533	0.358	0.363
002-01-16	153	0.035	0.020	0.051	0.793	0.554	0.029	0.691	0.504	0.194	0.288
002-11-16	496	0.062	0.122	0.178	0.928	0.736	0.012	0.477	0.456	0.452	0.170
003-07	2808	0.151	0.010	0.159	0.679	0.466	0.133	0.593	0.485	0.209	0.199
004-36	321	0.165	0.177	0.263	0.838	0.677	0.071	0.591	0.535	0.342	0.174
004-26-36	2292	0.005	0.180	0.181	0.521	0.421	0.424	0.337	0.233	0.647	0.226
101-15-28	606	0.001	0.188	0.188	0.125	0.043	0.852	0.120	0.059	0.863	0.224
101-22	1896	0.025	0.150	0.172	0.745	0.679	0.185	0.618	0.574	0.314	0.564
103-23	272	0.135	0.000	0.135	NA	NA	NA	NA	NA	NA	0.051
103-03	2191	0.152	0.267	0.348	0.811	0.629	0.069	0.598	0.494	0.294	0.251
103-42	195	0.429	0.786	0.913	0.929	0.832	0.001	0.914	0.579	0.001	0.098
103-29	56	0.651	0.699	0.874	0.737	0.708	0.236	0.546	0.298	0.249	0.176
104-19	2808	0.204	0.000	0.204	0.805	0.643	0.130	0.507	0.462	0.452	0.254
105-38	2808	0.271	0.225	0.380	0.803	0.663	0.138	0.623	0.585	0.293	0.249
106-06-39	2808	0.293	0.080	0.326	0.785	0.662	0.158	0.578	0.536	0.361	0.219
201-31	2808	0.353	0.288	0.427	0.847	0.753	0.124	0.624	0.571	0.286	0.163
202-05-33	457	0.000	0.195	0.195	0.953	0.869	0.000	0.830	0.495	0.010	0.257
203-09	1550	0.197	0.142	0.249	0.662	0.343	0.042	0.735	0.298	0.060	0.327
203-08-14	517	0.730	0.751	0.930	0.703	0.491	0.112	0.182	0.136	0.529	0.235
203-12-40	147	0.021	0.245	0.246	0.752	0.557	0.226	0.620	0.514	0.312	0.013
204-18	1048	0.056	0.370	0.381	0.574	0.364	0.340	0.517	0.388	0.341	0.207
204-10	1759	0.215	0.423	0.477	0.890	0.648	0.000	0.762	0.565	0.084	0.115
Unknown	NA	0.141	0.212	0.296	0.716	0.507	0.185	0.580	0.401	0.286	0.210
All		**0.183**	**0.196**	**0.308**	**0.758**	**0.587**	**0.156**	**0.584**	**0.470**	**0.306**	**0.226**

Table 2.5: Occupant specific summary of windows, blinds and lights status: combination of occupants, duration of observations, observed proportion of occupied time with windows open, observed mean unshaded fraction, proportion of occupied time with blinds fully lowered or fully raised, observed proportion of occupied time with lights on



(a) Distribution of indoor temperature by occupants (b) Distribution of indoor illuminance by occupants

Figure 2.7: Occupant specific distributions of indoor temperature and illuminance



(a) Day of the year (b) Indoor temperature (c) Indoor illuminance

Figure 2.8: Histograms of all observed values during occupied periods (gray rectangles) and at electronic questionnaire prompts (black lines)

Ref. occ.	Sex	Age	N_{ans}	Repr.	Ref. occ.	Sex	Age	N_{ans}	Repr.
1	M	20-30	51	No	19	M	40-50	248	Yes
2	M	30-40	127	No	21	M	20-30	112	No
3	F	20-30	243	Yes	25	F	20-30	138	No
4	M	20-30	244	Yes	26	F	30-40	313	Yes
5	F	20-30	120	No	27	F	20-30	177	Yes
6	F	50-65	578	Yes	28	M	20-30	214	Yes
7	M	50-65	202	Yes	30	F	20-30	410	Yes
9	M	50-65	222	Yes	31	M	50-65	661	Yes
10	M	30-40	263	Yes	33	F	30-40	37	No
11	M	20-30	384	Yes	34	M	20-30	126	No
15	M	30-40	103	Yes	36	M	40-50	276	Yes
16	M	20-30	415	Yes	37	M	20-30	283	Yes
17	M	20-30	47	No	38	F	50-65	224	Yes
18	M	50-65	204	Yes	39	F	50-65	429	Yes

Table 2.6: Summary of the surveyed occupants: Occupant reference number, sex, age category, number of answers and individual representativeness of their answers

conditions (Table 2.6), because the total duration of their employment within the laboratory was too limited. Therefore specific summaries of their answers may be biased.

Chapter 3

Behavioural modelling

Qu'est-ce que l'homme dans la nature ? Un néant
à l'égard de l'infini, un tout à l'égard du néant,
un milieu entre rien et tout.
*For after all what is man in nature? A nothing
in relation to infinity, all in relation to nothing,
a central point between nothing and all.*
Blaise Pascal (1623-1662), Pensées

This chapter begins with a discussion of approaches to the modelling of human behaviour and justifies the choice of probabilistic models in this context. We go on to present a detailed account of the mathematical methods for behavioural modelling, including the founding principles for model inference, the selection of key variables, the assessment of model quality and validation procedures.

3.1 Determinism and probabilistic models

All human actions have one or more of these seven causes: chance, nature, compulsion, habit, reason, passion and desire.
Aristotle (384-322 BC)

3.1.1 Some thoughts

Predicting human behaviour may seem at first glance to be an impossible task. Everyday our fellow humans surprise us by their sometimes irrational, inexplicable and unexpected actions.

It would be vain to expect a single individual to systematically reproduce the same clearly identified behavioural patterns, and so we may immediately discard simplistic approaches of the kind *"if stimulus A then action B"*. Rather our significant challenge is to understand and ultimately assess the consequences of humans' behavioural complexity.

A precise prediction of human actions is then out of reach, and we must cope with imperfect and approximate choices, where we at best understand and assess the great uncertainties around possible mainstream patterns. In this we are helped by the observation that people often reproduce some known patterns, as some actions are motivated by causes that we can identify. Based on such observations, we may hope to offer partial and limited predictions of the actions of interest to us, bearing in mind that we should avoid excessive generalisation.

Modelling human actions using a probabilistic approach allows for a more honest account of our observation-based knowledge, as the apparent contradiction of fluctuating choices performed in similar conditions is not an obstacle in this context. Such a probabilistic approach can be considered either as a consequence of an inside uncertainty or as a surrogate for our ignorance of unreachable but existing causes. In this latter case, we do not necessarily attempt to contradict the advocate of determinism Pierre-Simon de Laplace (1749-1827), who expressed in the introduction of the *Théorie analytique des probabilités* written in 1814:

> Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'Analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux.
>
> *"We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a given moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."*

We will rather admit that, in the present task to model human behaviour in building simulation, we stand far away from the knowledge of such a hypothetical intellect. In order to account for our unavoidable ignorance of the details of the system, non-deterministic approaches may enable us to contain the causes that we ignore or that are simply not of interest.

In conjunction with subject-matter considerations, probabilistic models have the possibility to identify associations between a given behaviour and an influencing variable, which will provide partly explained and partly unexplained variation. This latter part can then be attributed either to intrinsic variability or to our ignorance of further driving causes. We obtain thus a constructive separation between known systematic behavioural patterns and random variations that we will attempt to minimise.

3.1.2 Probabilistic models in building simulation

We do not arrive in a building or leave it at precisely determined schedules, neither do we systematically open windows or act on shading devices at precisely defined thresholds of temperature or at a given sun position. And yet this is how occupants' behaviour is currently represented in building simulation programs: not only does this crudely oversimplify reality, but it also neglects the study of the variability caused by the diversity in human behaviour.

To address this we must first perform observations revealing the intrinsic trends for humans to perform actions in the presence of driving stimuli. Our next step is to isolate a sufficient set of parameters that genuinely influence observed actions, and to infer a model formulation which adequately describes these relationships. In this we strive to achieve a compromise between complexity and usability. Finally, a robust validation procedure must assess the predictive power of the model.

3.2 Generalised linear models and logistic regression

3.2.1 Theoretical background

When modelling human actions, it is of interest to determine whether an action has taken place, and whether this action was influenced by one or more system variables. Formally, this implies the inference of a relationship between a dichotomous outcome variable Y and a set of p independent predictors $\mathbf{x} = (x_1, \ldots, x_p)$. We assume that the observations y_i of the random variable Y take values 0 or 1 for a negative or positive outcome, such as failure/success or no action/action.

Typically, the outcome variable Y is not uniquely determined by the value of $\mathbf{x}$, so the quantity of interest is the mean value of the outcome variable, given the values of the set of independent variables $\mathbf{x} = (x_1, \ldots, x_p)$. This quantity is called the *conditional mean* $E(Y|\mathbf{x})$ and we will set $p(\mathbf{x}) = E(Y|\mathbf{x})$ to simplify notation; by definition, $0 \leq p(\mathbf{x}) \leq 1$.

In the context of a classical linear model we would define a *linear predictor* $\eta = \mathbf{x}^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ and assume that $p(\mathbf{x}) = \eta$, but it would then be possible for p to take values outside the interval $[0, 1]$. It is thus necessary to use a suitable monotone transformation g of $p(\mathbf{x})$, satisfying $0 \leq g^{-1}(\eta) \leq 1$.

However, the classical least squares regression theory used for linear models is inappropriate for binary outcome variables. Firstly because the linear predictor η and the mean response are related by a transformation g of this latter, and also because of the violation of the crucial assumption that errors are normally distributed.

Generalised linear models

In order to overcome the above limitation, the class of *generalised linear models* (GLM) was developed, which extends the applicability of classical linear models beyond their restrictive conditions. A GLM requires the definition of:

- A *linear predictor* $\eta = \mathbf{x}^T\boldsymbol{\beta}$ expressing the effect of the predictors $\mathbf{x}$ and parameters $\boldsymbol{\beta}$ on the response;

- A *distribution function* f from the exponential family, which is the density of the outcome variable Y, defined as

$$f(y, \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right), \tag{3.1}$$

  where θ is a function of the linear predictor η and ϕ is called the dispersion parameter. Among many other, the normal, binomial, exponential and Weibull distributions are members of the exponential family [36]. For instance, if we set

$$y = \frac{k}{n}, \quad \phi = \frac{1}{n}, \quad \theta = \log\left(\frac{p}{1-p}\right), \quad b(\theta) = \log(1 + e^{\theta}), \quad c(y, \phi) = \log\begin{pmatrix} n \\ k \end{pmatrix},$$

then f is the *binomial density*,

$$f(n,k,p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 < p < 1, \quad k = 0,\ldots,n. \tag{3.2}$$

- A monotone *link function* g, providing the relationship between the linear predictor and the mean of the response: $\eta = g(\mu)$, with $\mu = E(Y)$.

In the case of a dichotomous outcome variable, f is taken as the binomial distribution, defining the binomial family of GLMs. Three common choices for link functions in this case are:

- The canonical logit link: $g(p) = \log\big(p(\mathbf{x})/(1-p(\mathbf{x}))\big)$,
- The probit[1] link: $g(p) = \Phi^{-1}(p)$, the quantile of the normal distribution,
- The complementary log-log link: $g(p) = \log(-\log(1-p))$.

The logit transformation is the *canonical link function* for binomial GLMs: it has the property $\eta = g(p) = \theta$, the canonical parameter in Equation 3.1. Furthermore, it has several desirable properties such as symmetry and direct interpretability of the obtained coefficients, as $\exp(\beta_i)$ measures the increase in the *log-odds* $\log\big(p/(1-p)\big)$ of the resulting probability, for a unit increase of x_i.

We will thus use the logit transformation for our models of the binomial family; such GLMs are often referred to as *logistic regression models*. In this case the linear predictor η is called the *logit* and the obtained probability function is called a *logistic function*, defined as

$$p(x_1,\ldots,x_p) = \frac{\exp(\mathbf{x}^T\boldsymbol{\beta})}{1+\exp(\mathbf{x}^T\boldsymbol{\beta})} = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}{1+\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}, \tag{3.3}$$

where $\mathbf{x}$ is the vector of the retained explanatory variables and $\boldsymbol{\beta}$ is the vector of the regression parameters. We will often refer to the parameter β_0 as the intercept and β_i ($i \neq 0$) as the slope associated with the variable x_i. It follows that the residuals have a binomial distribution with a mean of zero and variance $p(\mathbf{x}) \cdot (1-p(\mathbf{x}))$.

Further discussion of generalised linear models is provided in Dobson [36], while Hosmer and Lemeshow [37] focus on logistic regression.

Remarks on the logistic function

The logistic function with a single variable x has some important properties. In the case $\beta_1 > 0$, we notice that $p(x)$ is monotonously increasing, $p(x) \to 0$ for small x, and $p(x) \to 1$ for large x (and conversely if $\beta_1 < 0$). There is a *characteristic value* $x_{50} = -\beta_0/\beta_1$ for which $p(x = x_{50}) = 0.5$, and the variation of p is proportional to β_1 at this point, as $\mathrm{d}p/\mathrm{d}x(x = x_{50}) = \beta_1/4$.

The curves obtained using the three presented link functions differ mainly in the tails, as can be seen in Figure 3.1.

[1] The term "probit" comes from a contraction of *probability unit*, after which the term "logit" was subsequently introduced.

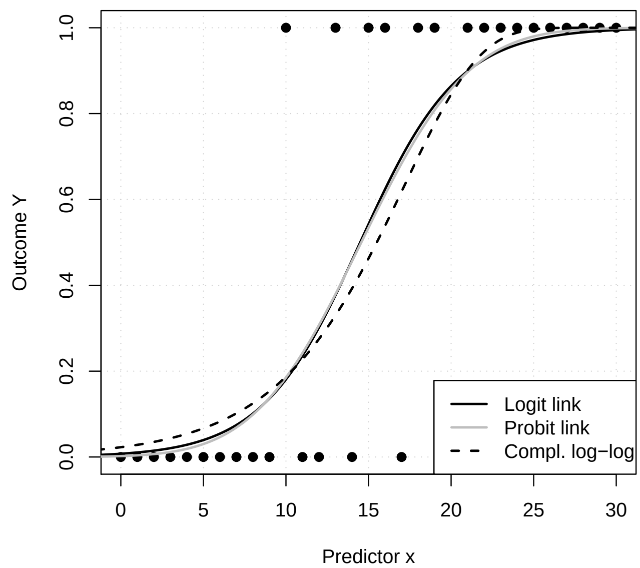


Figure 3.1: Comparison of the logit, probit and complementary log-log link functions obtained after fitting hypothetical binary data (solid points)

Maximum likelihood estimation

The classical least squares regression used for fitting linear models is not applicable to GLMs. We thus describe here briefly the appropriate method for fitting GLMs, known as *maximum likelihood estimation*.

Suppose that the sample observations y_i arise from a probability density function $f(Y|\boldsymbol{\beta})$ which is known, but for which the vector of parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is unknown. The *likelihood function* is the conditional probability of observing the sample, given $\boldsymbol{\beta}$, which is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}), \tag{3.4}$$

considered as a function of $\boldsymbol{\beta}$ for fixed y_i, assuming that the observations y_i arise from a random sample. The log-likelihood function is defined as

$$l(\boldsymbol{\beta}) = \log(L(\boldsymbol{\beta})) = \sum_{i=1}^{n} \log f(y_i|\boldsymbol{\beta}). \tag{3.5}$$

Differentiating the likelihood (or equivalently the log-likelihood) yields the value of $\boldsymbol{\beta}$ defined as the *maximum likelihood estimator* $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

In the case of a logistic regression model with a single explanatory variable x based on sample observations $(x_1, y_1), \ldots, (x_n, y_n)$, the likelihood function becomes

$$L(\beta) = \prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}. \tag{3.6}$$

We differentiate the log-likelihood $l(\beta)$ to find the value of β maximising $L(\beta)$, which leads to the estimated parameters $\hat{\beta}$. Except in special cases, the maximum likelihood estimate of β cannot be expressed explicitly. Algorithms used in statistical software to perform estimates of logistic regression parameters – such as non-iterative weighted least squares – are discussed by Hosmer and Lemeshow [37].

Standard errors and confidence intervals

The *observed information* $J(\boldsymbol{\beta})$ in a model with parameters $\boldsymbol{\beta}$ including p components and log-likelihood $l(\boldsymbol{\beta})$ and the *expected (or Fisher) information* $I(\boldsymbol{\beta})$ are defined as symmetric $p \times p$ matrices whose (i, j) elements are

$$J(\boldsymbol{\beta}) = -\frac{d^2 l(\boldsymbol{\beta})}{d\beta_i d\beta_j}, \qquad I(\boldsymbol{\beta}) = E(J(\boldsymbol{\beta})) = E\left(-\frac{d^2 l(\boldsymbol{\beta})}{d\beta_i d\beta_j}\right), \tag{3.7}$$

It can be shown (see for instance Davison [38]) that for a large sample size n, $\hat{\boldsymbol{\beta}}$ is distributed as a multivariate normal distribution $\mathcal{N}_p(\boldsymbol{\beta}^0, I(\boldsymbol{\beta}^0)^{-1})$, where $\boldsymbol{\beta}^0$ is the true value of $\boldsymbol{\beta}$.

As this result may be correctly approximated by $\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\theta^0, I(\hat{\boldsymbol{\beta}})^{-1})$, it follows that $I(\hat{\boldsymbol{\beta}})^{-1}$ is an estimator for the variance of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$. We may thus construct a $(1 - 2\alpha)$ level confidence interval for $\boldsymbol{\beta}^0$,

$$\left(\hat{\boldsymbol{\beta}} - z_{1-\alpha} I(\hat{\boldsymbol{\beta}})^{-1/2}, \hat{\boldsymbol{\beta}} - z_{\alpha} I(\hat{\boldsymbol{\beta}})^{-1/2}\right), \tag{3.8}$$

where $z_{1-\alpha}$ and z_α are the quantiles of the standard normal distribution.

3.2.2 Assessment of statistical significance

> Entia non sunt multiplicanda praeter necessitatem.
> *Entities should not be multiplied beyond necessity.*
> William of Ockham (ca. 1288-1347)

Likelihood ratio test statistic

We define the *deviance* of a generalised linear model with associated likelihood $L_{\text{fitted}}(\boldsymbol{\beta})$ as

$$D = -2\log\left(\frac{L_{\text{fitted}}(\boldsymbol{\beta})}{L_{\text{sat}}(\boldsymbol{\beta})}\right) = -2\log(L_{\text{fitted}}(\boldsymbol{\beta})), \tag{3.9}$$

with $L_{\text{sat}}(\boldsymbol{\beta})$ being the likelihood of a saturated model, which may be shown to be equal to one. The deviance may also be interpreted as equivalent to the residual sum-of-squares in linear regression.

We may now define a test for the significance of k independent variables added to a given model. Let us define one model without the k variables with likelihood L_0 and deviance D_0, and another model with L_1 and D_1. The *likelihood ratio test statistic* is defined as the deviance difference caused by the inclusion of the k predictors,

$$G = D_0 - D_1 = -2\log\left(\frac{L_0(\boldsymbol{\beta})}{L_1(\boldsymbol{\beta})}\right), \tag{3.10}$$

and asymptotically has a χ^2 distribution with k degrees of freedom, under the hypothesis that the model is correct. The associated p-value allows us to check the significance of adding new variables to the model.

Wald test

The Wald test for the significance of an independent variable is obtained by dividing the estimate of the obtained intercept or slope parameter $\hat{\beta}_i$ by its standard error $\mathrm{SE}(\hat{\beta}_i)$, giving $W = \hat{\beta}_i/\mathrm{SE}(\hat{\beta}_i)$. Under the null hypothesis, this ratio asymptotically follows a standard normal distribution, and the two tailed p-value $P(|z| > W)$ where $z \sim \mathcal{N}(0, 1)$ is easily obtained.

This test should not be used indiscriminately, as previous research conducted by Hauck and Donner [39] has shown that it often fails to reject the null hypothesis for significant predictors. The likelihood ratio test statistic is thus generally preferred for assessing the significance of individual predictors.

Information criteria

Other criteria have been developed to assess the statistical significance of predictors. For instance, Akaike Information Criterion (AIC) [40] and Bayesian Information Criterion (BIC) [41] propose expressions that balance the deviance of a model with a function of the number of predictors, in order to select an optimally parsimonious model. These are defined as

$$AIC = -2l(\hat{\boldsymbol{\beta}}) + 2p, \qquad BIC = -2l(\hat{\boldsymbol{\beta}}) + p \log n. \tag{3.11}$$

Following from the definitions of these criteria, the optimal model is associated with the lowest value of AIC or BIC.

3.2.3 Measures of goodness-of-fit

Statistical significance itself does not directly provide for an assessment of the quality of predictions given by a model. Furthermore, when a large database is used, a factor with tiny predicted effect may be statistically significant, but will not bring useful predictive improvements to future observations.

Several criteria have therefore been developed to determine how well a given distribution predicts the original outcome variable, i. e. the goodness of the fit of the model. In this paper we present the area under the ROC curve (AUC), the Nagelkerke R^2_N, the Somers' D_{xy} and the Brier score B. A comparison of goodness-of-fit tests is presented by Hosmer et al. [42].

Area under ROC curve

Having defined the probability of finding the outcome variable in the positive state $P(Y = 1|x)$, we use a cutpoint c such that $Y = 1$ if $P(Y = 1|x) > c$ and conversely for $P(Y = 1|x) \leq c$. Comparing values of Y predicted through this cutpoint with observed values of Y, four possibilities may arise: a predicted positive outcome is (i) truly positive, (ii) falsely positive (Type I error), a predicted negative outcome is (iii) truly negative, (iv) falsely negative (Type II error). We note the occurrences in these four exhaustive cases as respectively TP, FP, TN, FN. We may then define the *true positive rate* (or sensitivity, or hit rate) $TPR = TP/(TP+FN)$, the *false positive rate* (or fall-out) $FPR = FP/(FP+TN)$ and the *specificity* $SPC = 1 - FPR$.

We may plot the sensitivity against the complementary of the specificity for different values of the cutpoint between 0 and 1, giving the *receiver operating characteristic (ROC) curve* [43]. As the cutpoint varies, the intrinsic trade-off between sensitivity and specificity appears clearly. The area under the ROC curve will be called here the *AUC* index, and is a direct measure of the discriminating power of a given model. It may take values between 0.5 (we may as well toss a coin) and 1.0 (exact predictions), but values above 0.7 are generally considered to give acceptable discrimination, after Hosmer and Lemeshow [37].

In summary, this index allows a simple but direct comparison between predictions and observations.

Nagelkerke's generalised $\mathbf{R^2}$

Nagelkerke's generalised R^2 [44] measures the proportion of explained deviance in a model and is defined as

$$R_N^2 = \frac{1 - \exp(-D/n)}{1 - \exp(-D_{\text{null}}/n)}, \tag{3.12}$$

where D is the global log-likelihood ratio statistic of the considered model and $D_{\text{null}} = -2\log(L_{\text{null}}(\beta))$ refers to the null model, based on n observations. It extends the well-known definition of the proportion of explained variance R^2 used in linear models to the explained deviance in logistic models, although it does not exactly correspond to the proportion of response variation explained by the model. Values of R^2 obtained with this convention are generally much lower than their linear model counterparts.

Brier score

The Brier score measures the accuracy of a set of probability assessments. It is defined as the mean value of the squared difference between observed outcomes and their predicted probabilities:

$$B = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - y_i)^2, \tag{3.13}$$

where $\hat{p}_i$ are the predicted probabilities and y_i the corresponding observed values for the outcome variable, for each of n observations. It follows directly from the definition that a lower Brier score indicates a higher accuracy.

Somers' D_{xy} rank correlation

The Somers' D_{xy} parameter is defined as the difference between concordance and discordance probabilities. Given two individual values of the predictors x_0 and x_1, randomly sampled from the populations with outcome variable $Y_0 = 0$ and $Y_1 = 1$ respectively, we define then $D_{xy} = P(Y_1 > Y_0) - P(Y_0 > Y_1)$. Somers' D_{xy} may take values between 0 (random predictions) and 1 (perfect discrimination).

3.2.4 Variable selection procedure

In the procedure of inference of an optimal model, a central task is to correctly select an optimal set of variables to be retained. This should result in an accurate, yet parsimonious model, whose objective is to correctly predict future observations.

When too complex a model (one with too many parameters) is fitted, predictions may show lower agreement with further data, as some of the included parameters may contribute noise or express spurious correlations between the response and the predictors. This situation is known as *overfitting*. Our preferred model should thus include all the driving parameters that significantly influence the response to account for their impact and no more.

To this end, and based on the indicators introduced in Sections 3.2.2 and 3.2.3, we perform a *forward selection* where the best model with a single variable is first fitted, and proceed to consider models with further variables and assess their increased predictive value[2]. In this we first determine the best model containing two variables with their usual statistical indicators, and retain this additional variable depending on its significance and on the stability of the primary variable; continuing this procedure for other predictors until no added significance is obtained. Fit residuals (see Section 3.2.5) are checked at each step to assess the need for variable transformation or further refinements.

This done, a cross-validation procedure needs to be performed. This involves predicting the values of a part of the dataset (the validation set) using a model based on data from the complementary part (the training set). This allows for a direct unbiased assessment of the predictive power of a model.

Finally it is important to avoid the problem of *collinearity*, which arises when some of the explanatory variables are strongly correlated with each other. In this case, the solution for $\boldsymbol{\beta}$ will be unstable, that is some elements of $\boldsymbol{\beta}$ may display large variances, while marginal changes in the data can cause large changes in $\boldsymbol{\beta}$.

3.2.5 Model diagnostics

As with linear models, it is advisable to check whether some patterns are noticeable in the fit residuals, which can indicate violations of model assumptions or inadequate fit. There are several types of residuals used in logistic regression, amongst which we will examine two:

1. The Pearson residuals:
$$r_i = \frac{y_i - m_i\hat{p}_i}{\sqrt{m_i\hat{p}_i(1-\hat{p}_i)}} \tag{3.14}$$

2. The deviance residuals:
$$d_i = \mathrm{sign}(y_i - m_i\hat{p}_i) \cdot \sqrt{2\left(y_i \log\left(\frac{y_i}{m_i\hat{p}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i(1-\hat{p}_i)}\right)\right)} \tag{3.15}$$

where $\hat{p}_i$ are the fitted probabilities and y_i the number of times that $y = 1$ among the m_i repeats of x_i. Throughout this work we will systematically perform checks of the structure in the residuals, although this will be explicitly mentioned only when a particular problem is observed.

3.2.6 Ordinal logistic regression

We extend the previous discussion on logistic regression to the case where the outcome Y is not dichotomous but may take values in any of J categories. In this case, the associated distribution function is the *multinomial density* with denominator m and probability vector $\mathbf{p} = (p_1, \ldots, p_J)^T$, defined as

$$f(y_1, \ldots, y_J, n, p_1, \ldots, p_J) = \frac{n!}{y_1!y_2!\ldots y_J!}p_1^{y_1}p_2^{y_2}\ldots p_J^{y_J}, \qquad \text{with} \quad \sum_{i=1}^{J} y_i = n. \tag{3.16}$$

[2] In this work we will not use an automatic stepwise variable selection procedure, whose intrinsic problems are well documented (see for instance by Harrel [45] for a detailed discussion).

In the special case $J = 2$, Equation 3.16 reduces to the binomial distribution.

If there is a natural order among the J response categories, the situation can be simplified to a *proportional odds model*, where the probability for Y to fall in a category higher or equal to j is

$$p(Y \geq j|\mathbf{x}) = \frac{\exp(\alpha_j + \mathbf{x}^T\boldsymbol{\beta})}{1 + \exp(\alpha_j + \mathbf{x}^T\boldsymbol{\beta})}, \tag{3.17}$$

for $j = 1, \ldots, J$. With this formulation, the obtained regression parameters are consistent with the situation of a binary outcome, as for any value of j we obtain a binary logistic model for $p(Y \geq j)$. The distribution (3.17) includes J intercepts α_j and (common) regression parameters $\boldsymbol{\beta}$. The proportional odds model offers thus a parsimonious representation of ordinal responses. However the underlying assumption that the coefficients $\boldsymbol{\beta}$ are independent of j needs to be checked, by assessing how the variations of Y relate to the mean of $\mathbf{x}$.

Statistical significance may be examined in a similar way as for binary models; likewise tests for goodness-of-fit, discussed thoroughly by Fagerland et al. [46]. A detailed general discussion of ordinal logistic regression models is provided by Agresti [47].

3.3 Random processes

In modelling human behaviour, we are often interested in the description of the dynamic evolution of a random variable X with time t. Let then X_t denote a *random* or *stochastic process* taking values in the *state space* $\mathcal{S} = \{1, \ldots, S\}$. The process consists then in the set $\{X_{t_0} = s_0, X_{t_1} = s_1, \ldots, X_{t_n} = s_n\}$ which is the collection of the observed states of X for all values at time t. It is possible to consider t either as discrete – in this case X_t is a discrete-time random process – or continuous – X_t is then a continuous-time random process.

3.3.1 Bernoulli processes

A Bernoulli process is a discrete-time stochastic process consisting of a sequence of independent identically distributed Bernoulli random variables X_t, for $t = \{t_0, \ldots, t_n\}$, such that for each t, $X_t \in \{0, 1\}$, and $P(X_t = 1) = p$ for all t. If p is constant the process is *homogeneous* (or *stationary*) whereas if p varies with time the process is *inhomogeneous*. Put formally,

$$P(X_{t_{i+1}} = s_{i+1}|X_{t_i} = s_i, X_{t_{i-1}} = s_{i-1}, \ldots, X_{t_0} = s_0) = P(X_{t_{i+1}} = s_{i+1}) = p. \tag{3.18}$$

The independence assumption implies that past and present values of X_t do not provide any information about future outcomes. A Bernoulli process has thus no memory and each step may be seen as the start of a new process. For instance, it correctly describes a set of coin tosses (with $p = 1/2$ for an unbiased coin). This kind of process may be generalised to a larger set of possible outcomes, each with probability p_j which may or may not vary with time. However, in our context the assumption of independence is often excessively simplistic.

3.3.2 Markov processes

Markov processes extend Bernoulli processes by including the current state X_{t_i} to predict $X_{t_{i+1}}$. This assumes that all of the information describing the process history is carried by

X_{t_i}, which is expressed by the *Markov property*

$$\begin{aligned} P(X_{t_{i+1}} = s_{i+1} | X_{t_i} = s_i, X_{t_{i-1}} = s_{i-1}, \ldots, X_{t_0} = s_0) \\ = P(X_{t_{i+1}} = s_{i+1} | X_{t_i} = s_i) = p_{s_i, s_{i+1}} \end{aligned} \tag{3.19}$$

Markov models which predict states at discrete equally-spaced times are called *Markov chains*. In this case, we can define transition probabilities p_{ij} that can be arranged as a transition matrix P, with the properties that $\sum_i p_{ij} = 1$ and $p_{ij} \geq 0$, for all i, j. With this convention, it can be shown that the $(i,j)^{\text{th}}$ element of P^n is the n-step transition probability $p_{ij}^{(n)} = P(X_n = j | X_0 = i)$.

Markov chains, defined by Equation 3.19 are called first-order chains: they depend only on the current state of X_t through p_{ij}, while Bernoulli processes – with the independence assumption (Equation 3.18) – are zeroth-order chains, and we have $p_{ij} = p_j$. It is possible to extend this reasoning to chains of order $m > 1$, where transition probabilities depend on the m previous states:

$$\begin{aligned} P(X_{t_{i+1}} = s_{i+1} | X_{t_i} = s_i, X_{t_{i-1}} = s_{i-1}, \ldots, X_{t_0} = s_0) \\ = P(X_{t_{i+1}} = s_{i+1} | X_{t_i} = s_i, X_{t_{i-1}} = s_{i-1}, \ldots, X_{t_{i-m+1}} = s_{i-m+1}) = p_{s_{i-m+1} \cdots s_{i-1} s_i s_{i+1}}. \end{aligned} \tag{3.20}$$

In this case the significance of increasing the order of the chains should be tested. For instance, we may estimate $\hat{p}_{ij}$ and $\hat{p}_j$ from the observed counts of transitions n_{ij}. The likelihood ratio statistic may then be used to compare the zeroth and first order chains,

$$W = 2 \sum_{i,j} n_{ij} \log\left(\frac{\hat{p}_{ij}}{\hat{p}_j}\right), \tag{3.21}$$

where W is asymptotically chi-squared distributed with $(S-1)^2$ degrees of freedom. See Davison [38] or Norris [48] for further theoretical details and Avery and Hendersen [49] for an interesting application.

Markov processes may also be defined on a continuous-time basis (see for instance Norris [48] for further discussion) in which delays between transitions are exponentially distributed.

3.4 Survival analysis

Survival analysis typically attempts to model data in which it is the time elapsed until an event occurs that is of interest. This statistical method has long-since been applied in reliability studies and biomedical research, where the survival durations until a failure or a death is studied. In our context, this type of model may be relevant to predict the delays until actions are performed by building occupants, such as closing a window once it has been opened.

The durations t are modelled as a non-negative random variable T, with cumulative distribution function $F(t) = P(T \leq t)$ and corresponding probability density function $f(t)$. We define then the associated survival function (or reliability function) $S(t) = P(T > t) = 1 - F(t)$ and the hazard function $h(t) = f(t)/S(t)$. It follows from these definitions that $f(t) = -\mathrm{d}S(t)/\mathrm{d}t$ and $h(t) = \mathrm{d}\log(S(t))/\mathrm{d}t$.

A strength of these statistical methods lies in the possibility of including survival times containing partial information in the analysis, such as survival for at least a given period of time, without needing further knowledge about the process after this time. These non-comprehensive observations are denoted as *censored data*[3].

Non-parametric estimates

Provided r_i, the number of surviving elements until time t_i (including censored observations) and d_i, the number of "failures" at time t_i, we find that the conditional probability to survive beyond t_i, knowing that the subject is alive just before t_i is given by $(r_i - d_i)/r_i$. The survival function can then be estimated by the non-parametric *Kaplan-Meier* or *product-limit estimator* [50],

$$\hat{S}_{KM}(t) = \prod_{i|t_i<t} (1 - \frac{d_i}{r_i}) \tag{3.22}$$

which defines unambiguously an estimator of the distribution $F(t) = 1 - S(t)$ for durations.

Parametric estimates

However, in order to perform a parametric estimation, a survival distribution has to be fitted using maximum likelihood regression. We consider here two particular distributions to be used in our study:

- The *exponential distribution*, with $f(t) = \lambda \exp(-\lambda t)$, $S(t) = \exp(-\lambda t)$, which assumes a constant hazard rate $h(t) = \lambda = \text{const}$. This is the simplest model.
- The *Weibull distribution*, with $f(t) = \lambda\alpha(\lambda t)^{\alpha-1}\exp(-(\lambda t)^{\alpha})$, $S(t) = \exp(-(\lambda t)^{\alpha})$ and $h(t) = \lambda\alpha(\lambda t)^{\alpha-1}$, where the parameter α is called the *shape* and λ the *scale*. This offers more flexibility. In the special case $\alpha = 1$, it corresponds to the exponential distribution. When $\alpha < 1$, $h(t)$ decreases with t, and failures are more likely at small times; the converse is true for $\alpha > 1$.

As for generalised linear models (Section 3.2), maximum likelihood estimation is used to calculate the parameters of $S(t)$. Equation 3.4 has then to be modified to include censored data, which yields

$$L(\boldsymbol{\beta}) = \prod_{i=1\ (y_i \text{ uncens.})}^{n} f(y_i|\boldsymbol{\beta}) \cdot \prod_{i=1\ (y_i \text{ censored})}^{n} S(y_i|\boldsymbol{\beta}), \tag{3.23}$$

where y_i denotes either t_i if uncensored or the censoring time. Thus, Equation 3.4 is augmented by a contribution that is the probability that T_i is higher than the censoring time. We refer the reader to Davison [38] or Kleinbaum and Klein [51] for further details on survival analysis.

[3] For example, survival analysis could be used to model the duration for which an occupant will leave a window closed following their arrival, or open since its opening, according to relevant driving parameters. Observations such as closing on departure or the absence of opening during the whole occupancy period can be included in the modelling of duration of window openings as censored data.

Chapter 4

Modelling actions on windows

Pour examiner la vérité, il est besoin, une fois dans sa vie, de mettre toutes choses en doute autant qu'il se peut.
If you are to be a real seeker after truth, it is necessary that at least once in your life you doubt, as far as possible, all things.
René Descartes (1596-1650)

Based on almost eight years of continuous measurements (Section 2.1), we have analysed in detail the influence of occupancy patterns, indoor temperature and outdoor climate parameters (temperature, wind speed and direction, relative humidity and rainfall) on window opening and closing behaviour. In this we have also considered the variability of behaviours between individuals. This chapter begins with a detailed state of the art, followed by a presentation of some key findings from these analyses. We go on to develop and test several modelling approaches, including Bernoulli random processes based on probabilistic logistic models, Markov chains and continuous-time random processes. Informed by detailed statistical analysis and cross-validation of each variant, we propose a hybrid of these techniques which models stochastic window usage behaviour with optimal accuracy[1].

4.1 Introduction

We present in this section the current state of research relating to the prediction of occupants' actions on windows, and identify the key advances that have been made in this field together with some open questions.

4.1.1 State of the art

Pioneering investigations in residential buildings performed by Dick in 1951 [54], Brundrett from 1975 to 1979 [55, 56] and Lyberg in 1982 [57] reached agreement on the fact that actions on windows were positively correlated with external air temperature, and marginally negatively associated with wind speed. The influence of other stimuli was not examined.

[1] A substantial proportion of this chapter was presented at the 11^{th} International Building Performance Simulation Association Conference [52] where it was awarded the Arup Prize for the IBPSA Student Paper on Simulation and Design. A more complete paper was also published in the journal *Building and Environment* [53], where it received the Best Paper Award for 2009.

In 1984, Warren and Perkins [58] showed – using stepwise multiple correlation analysis – that external air temperatures accounted for 76% of the observed variance in the opening status of windows, the sunshine for an additional 8% and wind speed for 4%. A linear relationship between the percentage of rooms with at least one window open (a boolean response) was linked to the external temperature. However this was not formally equivalent to a probability, as p may take values outside of the range $[0, 1]$. A questionnaire conducted as a part of this study also revealed for the first time that occupants act on their windows particularly often on arrival and at departure.

A first attempt to develop a coherent mathematical model to predict the state of windows was performed by Fritsch et al. in 1991 [59, 60]. Based on measurements of the opening angle of four windows in four office rooms recorded every half an hour in the LESO building (see Section 2.1.1), a discrete-time Markov process (see Section 3.3.2) was developed to predict transitions between bins of opening angles. The model is formulated as Markov chains defining transition probabilities between six states, each corresponding to a definite class of opening angles and adjusted for four different outdoor temperature ranges [2]. The model includes transition matrices adapted to each occupant, in order to account for significant observed variations. The dependance of the percentage of opened windows versus wind speed and sunshine was also examined, but no significant variation was observed for wind speeds lower than 5-6 m/s. Although south facing vertical irradiance was observed to be correlated, especially in the mid-season, only outdoor temperature was retained as a model parameter.

Towards the end of the 1990s, interest in the adaptive approach to thermal comfort drew attention to the relationship between behaviour and thermal satisfaction. This led to several measurement campaigns in Pakistan from 1993 to 1996 [62, 63], the United Kingdom [64, 65] and five European countries [66].

Based on measurements from these three surveys, Nicol [67] proposed in 2001 the first coherent probability distribution for the prediction of the state of windows, as logistic functions (see Section 3.2) of indoor and outdoor temperature. In most cases the correlation with indoor temperature is similar to that with outdoor temperature, but Nicol recommends the use of outdoor temperature on the basis that it is an input of any simulation program, while indoor temperature is an output. However, Nicol and Humphreys [68] later reported in 2004 that indoor temperature was a more coherent predictor for the use of windows than outdoor temperature. This approach may seem more sensible: as Robinson [69] points out, predicted probabilities of interaction are otherwise independent of the design of the buildings in which occupants are accommodated.

Rijal et al. [70, 71] subsequently published in 2007 a more refined model, considering both indoor and outdoor temperature. A logistic model with two variables was derived for the probability of a window to be open. A deadband of ± 2°C for θ_{in} and ± 5°C for θ_{out} was defined to distinguish the probability of opening from that of closing. This refinement, although not based on any observed actual openings and closings, potentially solves the problem of repeated actions that would take place if a single distribution was used. The implementation in ESP-r of what has become named the *Humphreys algorithm*, involves the following steps:

1. Inputs parameters θ_{out}, $\theta_{\text{out,rm}}$, θ_{in} and θ_{op} are retrieved;

[2] These six classes are defined by the thresholds $0°, 1°, 15°, 35°, 60°, 90°$. The choice of narrower classes at small angles is based on the desire to obtain meaningful airflow rates, which vary more sharply at low angles (see for instance [61]). The four outdoor temperature domains are defined by the thresholds 0°C, 8°C and 16°C.

2. The comfort temperature θ_{comf} is computed from $\theta_{\text{out,rm}}$ according to the adaptive algorithm included in the CEN standard (Section 6.1.1).

3. If $|\theta_{\text{op}} - \theta_{\text{comf}}| > 2°\text{C}$, the probability of action is calculated by $\text{logit}(p) = 0.171 \cdot \theta_{\text{op}} + 0.166 \cdot \theta_{\text{out}} - 6.4$; else, if $|\theta_{\text{op}} - \theta_{\text{comf}}| \leq 2°\text{C}$ the window state remains unchanged.

4. This probability is compared with a random number to determine the next window state.

A comparison between observed and simulated window opening proportions for several indoor and outdoor temperatures ranges is provided as validation.

Page developed in 2006 [72] a behavioural model of actions on windows, relating indoor temperature and pollutant concentration with interaction probability, based on thresholds defined by Fanger's thermal (see Section 6.1.1) and aeraulic [73] comfort models. Interaction occurs when the calculated indoor concentration exceeds a critical concentration, or when hot or cold comfort limits (defined by $|PMV| > 2$) are surpassed.

Based on their summer field survey, Haldi and Robinson [74] suggest that in summer the strong correlation between indoor and outdoor conditions in naturally-ventilated buildings could dampen the efficiency of logistic models with these two variables. The works of Yun and Steemers [75, 76] seem to strengthen this hypothesis. Rijal et al. [77] have subsequently published a refinement of the Humphreys algorithm, including a window opening effectiveness parameter. This modification imposes a window to be closed if $\theta_{\text{out,rm}} > 28.1°\text{C}$ and $\theta_{\text{out}} > \theta_{\text{in}} + 5°\text{C}$.

Yun and Steemers [75, 76] performed a field survey on 6 offices facing east, west or south in 2 buildings, during 3 months in summer only. Indoor temperature was retained as a driving stimulus, considering that *"the prediction as a function of external temperatures cannot be considered as an intrinsic result"*, in agreement with Robinson's observation [69]. It was noticed that changes in window states mainly occurred on arrival or at departure.

A useful feature of [75, 76] is the use of separate probabilistic sub-models for the opening of windows at arrival and during occupancy[3]. Retained offices did not enable night ventilation, so actions on departure are not considered (windows are assumed to be closed at departure). Furthermore, this model predicts changes in window state from open to closed and from closed to open using indoor temperature and previous window state as predictors. Analyses show that the addition of outdoor temperature is not significant, in agreement with the summer survey of Haldi and Robinson [74]. The probability of opening on arrival is a logistic model based on indoor temperature, while a linear function is proposed for actions during occupancy. A short comparison between participants is also proposed. The final model, which has been implemented in ESP-r [80], retains thus indoor temperature, occupancy transitions and previous window state. Whilst this model may describe actions during the summer (for cases where deliberate night ventilation is not exercised), there is a question as to its validity during the winter period.

Herkel et al. [81, 82] also pointed out that most window openings can be associated with the arrival of an occupant, and so proposed separate sub-models for windows openings and closings on arrival, at departure and during occupancy. These sub-models consider outdoor temperature as the driving stimulus, based on the observation that this variable had a higher correlation with the hourly mean value of opening status of the monitored windows[4].

[3] Note that this distinction is not new, as it has been previously employed to the problem of light switching prediction [78, 79].

[4] An additional effect from season was noticed (eg. similar outdoor temperatures do not imply same action probabilities in spring or autumn), although this is not considered in the final model.

The final model is then formulated as six probabilities of opening and closing for arrival, intermediate presence and departure, given as quadratic functions of outdoor temperature. Such a distribution fits data acceptably well on observed small action probabilities, but is not formally consistent as p may take values outside of the range $[0, 1]$.

Based on eight months of measurements in six offices Mahdavi et al. [83, 84] noticed that windows are opened more frequently early in the day, after lunch time and towards the end of working hours; while closing actions are observed more frequently before occupants leave their offices for the day. An increased proportion of open windows was noticed when outdoor temperature rises to 26°C, with a decreasing trend above this value. There was no attempt to infer a predictive model.

Supported by two surveys (once in summer and once in winter), each consisting of a single questionnaire concerning window opening behaviour sent to 4948 dwellings in 2008, Andersen et al. [85] used multiple logistic regression analysis to deduce odd ratios for the significance of a set of variables. It was noticed that the respondent's gender, the outdoor temperature, the perceived illumination, air quality and noise levels had a statistically significant impact on "perceived" window opening behaviour. However, these conclusions were not based on any direct measurements.

4.1.2 Key advances, open questions and research needs

From the above review, we conclude the following:

- Explored methods for the simulation of actions on windows include logistic models and discrete-time Markov processes.

- Thermal stimuli have been shown to be the predominant causes for actions (indeed non-thermal variables are generally ignored), but no clear consensus is reached whether indoor or outdoor temperature should be used as the independent variable in the simulation of actions on windows.

- It is known that the influence of occupancy patterns is important.

- Independent studies have observed specificities in summer behaviour, but seasonal variations in behaviour have yet to be taken into account.

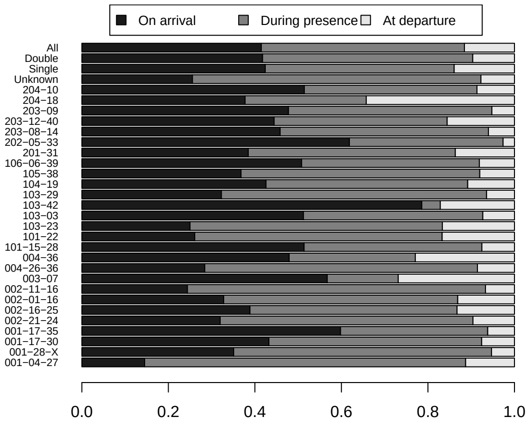
- The treatment of occupant behaviour towards night ventilation is not considered.

- Window opening angles are mostly ignored, even though these are crucial for reliable air flow prediction.

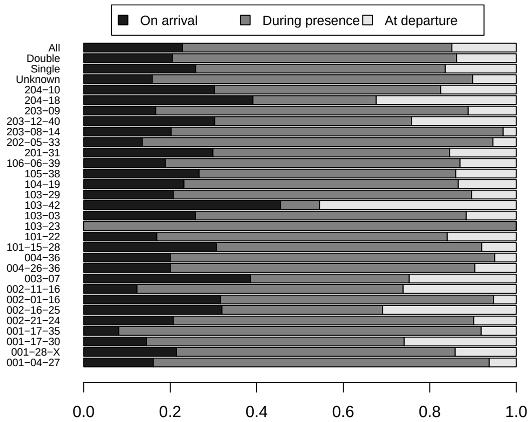
- Published studies do not provide any common robust cross-validation procedure, which prevents any comparison of quality between published models.

- The case of offices with several occupants is not specifically treated (authoritarian versus democratic behaviour).

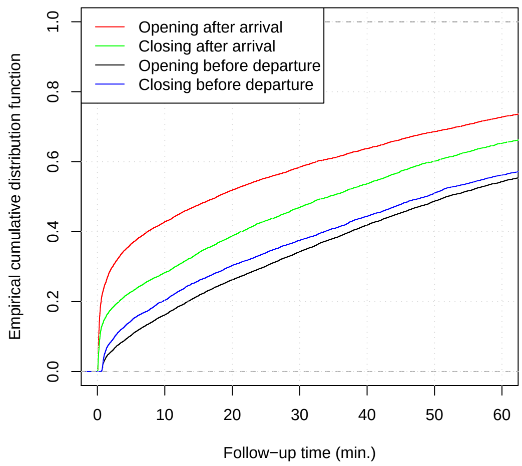
- Existing models are informed by measurements in office buildings and behaviour in residential environment is not specifically treated.



(a) Occupancy-dependent repartition of window openings for each combination of occupants and for all single and all double occupied offices



(b) Occupancy-dependent repartition of window closings for each combination of occupants and for all single and all double occupied offices



(c) ECDF of time intervals between occupancy transitions and actions on windows

Figure 4.1: Observed proportions of window openings and closings taking place in different occupancy situations (dark grey: arrival, grey: during occupancy, light grey: departure)

Actions	Arrival	Intermediate	Departure	Total
Opening	1905 (1.9%)	2923 (0.27%)	266 (0.27%)	5094 (0.40%)
Closing	713 (0.7%)	3566 (0.33%)	794 (0.81%)	5073 (0.40%)
Left open	25424 (25.8%)	356707 (32.85%)	25892 (26.33%)	408023 (31.81%)
Left closed	70313 (71.5%)	722762 (66.56%)	71403 (72.60%)	864478 (67.39%)
Total	98355 (7.7%)	1085958 (82.6%)	98355 (7.7%)	1282668 (100%)

Table 4.1: Classification of observed actions on windows based on 5 minute time steps with respect to occupancy status

4.2 Results

We present in this section some preliminary observations on window opening behaviour followed by results obtained from the application of the different models inferred on the basis of the concepts introduced in Chapter 3. The statistical software R [86] and statistical analysis functions of its package Design [87] were used for all data analyses and for programming the models.

4.2.1 Preliminary observations

Occupancy patterns. On average, occupants performed 2.409 opening (and closing) actions per day. As noticed in previous surveys [58, 76, 82], actions on windows more often occur when occupants arrive or leave their offices (Figure 4.1). From the figures presented in Table 4.1, we deduce that 37.4% of all recorded opening actions take place on arrival (5.2% at departure), and that 14.1% of closing actions happen on arrival (15.7% of them at departure). For this a first difficulty is the arbitrary choice of a temporal threshold for the definition of events occurring just after arrival or before departure. We plot the empirical cumulative distribution functions (ECDF) of time intervals between occupancy transitions and actions on windows (Figure 4.1(c)), and we notice that a threshold of five minutes defines a limit above which the slopes of the ECDFs remain relatively constant, suggesting that all events related to occupants' arrival and departure have by that time been captured. A possible explanation is that during this time occupants may perceive considerable differences in thermal and/or olfactory stimuli compared to their previous (possibly external) environment and their offices. These differences may be exacerbated by the absence of actions while the occupant was not present.

For offices with two occupants, we notice a slightly greater proportion of openings and closings during occupancy (Figure 4.1(a)). However, in this case many actions on arrival or departure are classified as intermediate actions, for instance when an occupant arrives in an already occupied office and opens the window. This question relates to the issue of group actions, discussed in Section 4.2.5.

Influence of environmental parameters. Our first step in the inference of a model linking the occurrence of open windows and one or several physical predictors involves examining the observed proportion of windows open with respect to the measured physical parameters. In Figure 4.2 we chart separately these proportions (based on observations taken every 10 minutes) grouped by bins of each measured physical parameter, with their statistical uncertainties[5].

A clearly increasing proportion may be observed for θ_{in} rising from 20°C to 28°C, with a possibly significant decrease above this range[6] (Fig. 4.2(a)). A similar phenomenon occurs with θ_{out}, the maximum proportion being reached around 26°C, above which a decrease is clearly significant (Fig. 4.2(b)). This type of behaviour was previously remarked by Rijal et al. [77]. Both thermal variables are thus clearly linked with actions on windows, in agreement with previous research. A less sharp decrease in the proportion of windows open is visible when outdoor humidity rises (Fig. 4.2(c)). Further examination will assess whether these variations are intrinsically influenced by each parameter, as these variables are correlated: $\rho(\theta_{\text{in}}, \theta_{\text{out}}) = 0.62$, $\rho(\theta_{\text{out}}, \phi_{\text{out}}) = -0.45$.

Mean wind speed v_{wind} (Fig. 4.2(e)) increase is linked with a decreased proportion of open windows for $v_{\text{wind}} > 2$ m/s. No particular variation may be observed with wind direction (Fig. 4.2(f)). No clear pattern may be noticed with respect to rainfall D_{prec} (Fig. 4.2(d)), which is slightly correlated with relative humidity: $\rho(D_{prec}, \phi_{\text{out}}) = 0.25$.

Based on these preliminary examinations, a relevant model would include in order of priority θ_{in}, θ_{out} and possibly ϕ_{out} and v_{wind}.

Variability between occupants and personal patterns related to actions on windows. We have noticed that, while the climatic conditions are fairly similar between the studied offices, some occupants use their windows more frequently than do others (see Table 2.5). Reinhart [79] suggested a distinction between "active" and "passive" occupants in the case of actions on blinds and lights; perhaps a similar classification may be applied for windows? To test this hypothesis we shall first produce generalist models based on all occupants, and go on to develop a method to treat variability among occupants in Section 4.2.5.

Towards an appropriate model for the prediction of actions on windows. These initial observations enable us to draw some first conclusions about the general form of a comprehensive model:

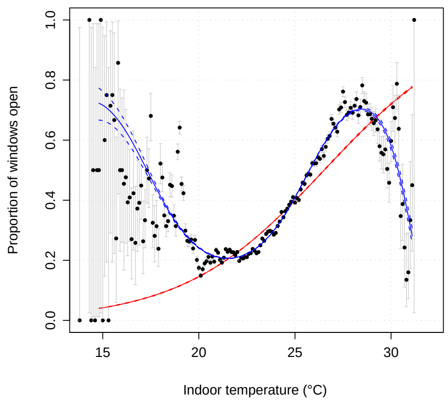
- Occupancy patterns should be integrated, for instance to facilitate separate treatment of actions on arrival, departure and during occupancy. This implies the necessity to explicitly model occupancy itself;

- Among possible other parameters, indoor and outdoor temperatures are the main driving stimuli for actions on windows;

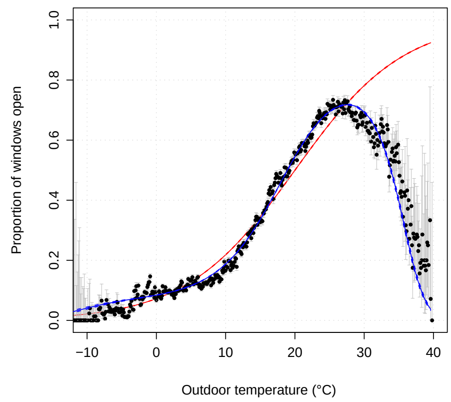
- A possible refinement could be to distinguish between "active" and "passive" users, and possibly between single and multi-occupied offices.

[5] We use the binomial cumulative distribution function to compute exact confidence intervals, rather than the more usual asymptotic normal interval. See Brown et al. [88] for further discussion on these calculations.

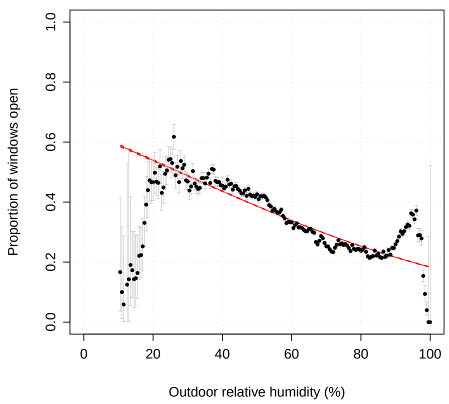
[6] A higher proportion of windows open at low θ_{in} is noticeable in Figure 4.2(a). This is caused by a window left open while θ_{out} is low. The 95% confidence intervals are large, owing to the very small number of observations within the relevant bins.



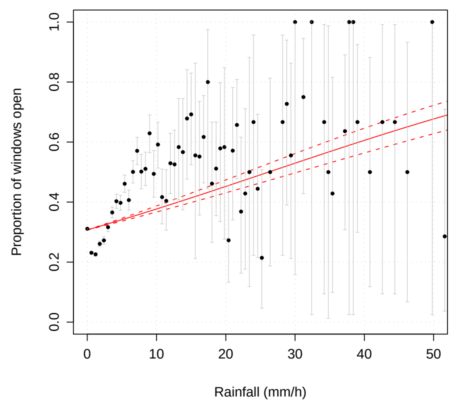
(a) Indoor temperature, bin width 0.1°C



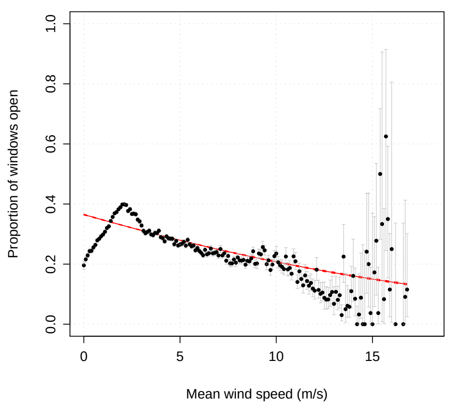
(b) Outdoor temperature, bin width 0.1°C



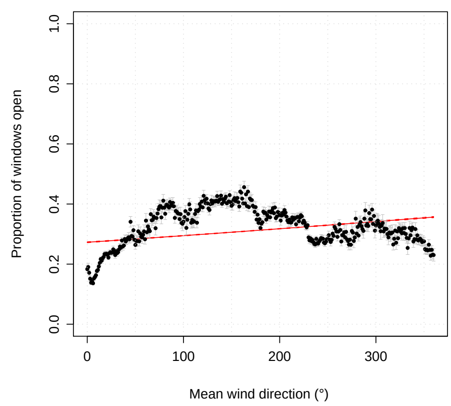
(c) Outdoor humidity, bin width 0.5%



(d) Rainfall, bin width 0.5 mm/h

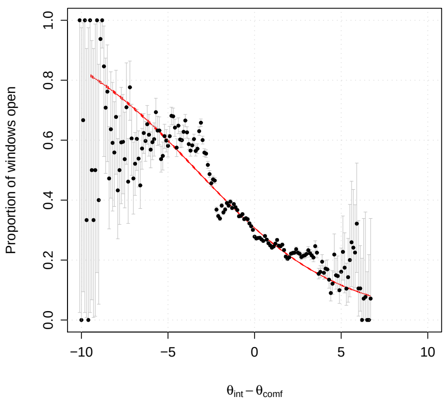


(e) Mean wind speed, bin width 0.1 m/s

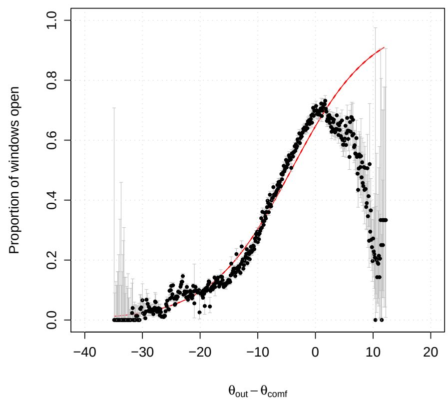


(f) Wind direction, bin width 1°

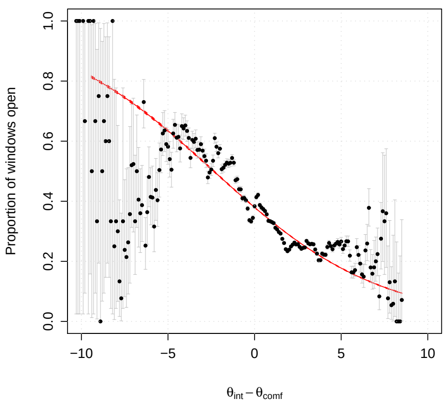
Figure 4.2: Observed proportion of windows open for specified bandwidths as a function of different measured physical parameters with binomial 95% level confidence intervals and fitted logistic regression curves (red: linear, blue: fourth degree polynomial)



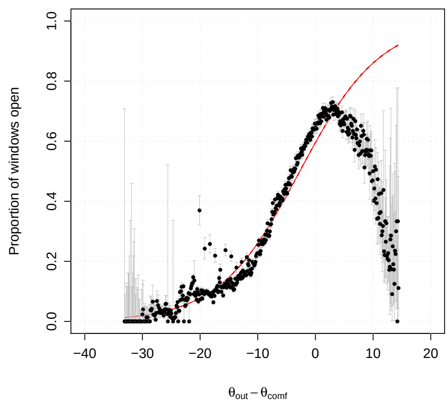
(a) $\theta_{\mathrm{in}} - \theta_{\mathrm{in,comf}}^{(\mathrm{CEN})}$, bin width 0.1°C



(b) $\theta_{\mathrm{out}} - \theta_{\mathrm{out,comf}}^{(\mathrm{CEN})}$, bin width 0.1°C

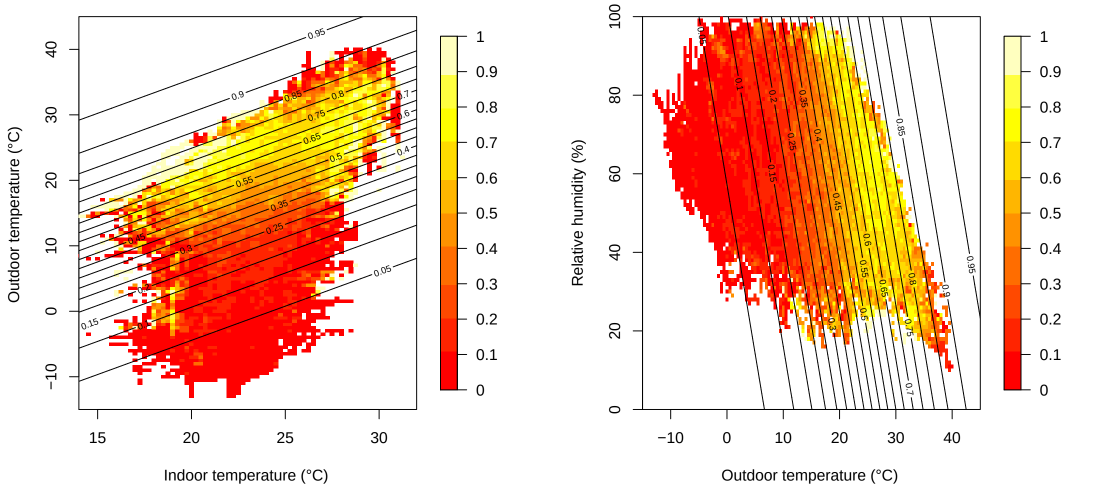


(c) $\theta_{\mathrm{in}} - \theta_{\mathrm{in,comf}}^{(\mathrm{ASHRAE})}$, bin width 0.1°C



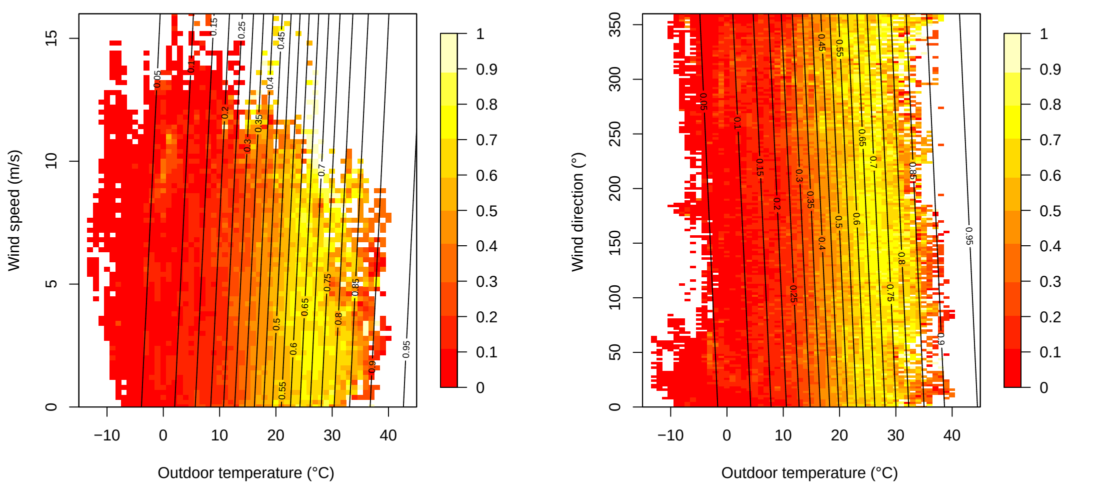
(d) $\theta_{\mathrm{in}} - \theta_{\mathrm{out,comf}}^{(\mathrm{ASHRAE})}$, bin width 0.1°C

Figure 4.3: Observed proportion of windows open for specified bandwidths as a function of deviations from comfort temperatures with binomial 95% level confidence intervals and fitted logistic regression curves



(a) Outdoor and indoor temperatures, bin widths 0.5°C

(b) Outdoor temperature and relative humidity, bin widths 0.5°C and 1%RH



(c) Outdoor temperature and wind speed, bin widths 0.5°C and 0.2 m/s

(d) Outdoor temperature and wind direction, bin widths 0.5°C and 1°

Figure 4.4: Observed proportion of windows open for specified bandwidths of different measured physical parameters with contour lines of equal fitted probabilities curves

Variable	a	Wald Z	b	Wald Z
θ_{out}	-2.5506 ± 0.0043	-594.0	0.12750 ± 0.00025	502.1
θ_{in}	-7.202 ± 0.025	-288.0	0.2716 ± 0.0011	257.7
ϕ_{out}	0.5693 ± 0.0078	72.9	-0.02063 ± 0.00012	-179.0
v_{wind}	-0.5542 ± 0.0030	-185.9	-0.07857 ± 0.00077	-102.7
α_{wind}	-0.9809 ± 0.0032	-304.3	0.001084 ± 0.000017	64.4
D_{prec}	-0.8126 ± 0.0018	-460.1	0.0311 ± 0.0022	14.0
f_R	-0.7944 ± 0.0018	-440.9	-0.2395 ± 0.0075	-32.1
f_{WS} (< 1.5)	-0.8731 ± 0.0037	-236.9		
f_{WS} (1.5-2.5)			0.3940 ± 0.0050	79.5
f_{WS} (2.5-4.7)			0.0958 ± 0.0050	19.2
f_{WS} (> 4.7)			-0.2857 ± 0.0053	-54.2
f_{WD} (North)	-1.2096 ± 0.0034	-358.7		
f_{WD} (East)			0.6589 ± 0.0053	124.8
f_{WD} (South)			0.6979 ± 0.0046	150.4
f_{WD} (West)			0.3445 ± 0.0050	69.2
$\theta_{\text{out}} - \theta_{\text{comf}}^{(\text{ASHRAE})}$	0.3866 ± 0.0029	134.1	0.14290 ± 0.00029	486.1
$\theta_{\text{in}} - \theta_{\text{comf}}^{(\text{ASHRAE})}$	-0.4908 ± 0.0023	-215.0	-0.2085 ± 0.0010	-206.6
$\theta_{\text{out}} - \theta_{\text{comf}}^{(\text{CEN})}$	0.6031 ± 0.0032	188.0	0.14310 ± 0.00029	485.6
$\theta_{\text{in}} - \theta_{\text{comf}}^{(\text{CEN})}$	-0.8145 ± 0.0018	-457.4	-0.2426 ± 0.0011	-214.1

Table 4.2: Regression parameters for logistic models including a single variable (in all cases, $p < 0.001$ according to Wald and likelihood ratio tests). For models including dummy variables with several levels (f_{WS} and f_{WD}), fitted values for baseline levels are given as intercepts.

4.2.2 Models based on single probability distributions

From now on we use the following notation for all the models based on logistic regression:

$$\begin{aligned} \text{logit}(p) = \log(\frac{p}{1-p}) &= a + b_{\text{in}}\theta_{\text{in}} + b_{\text{out}}\theta_{\text{out}} \\ &+ b_\phi \phi_{\text{out}} + b_R f_R + b_{WS} f_{WS} + b_{WD} f_{WD}, \end{aligned} \tag{4.1}$$

where a and b_i are the regression parameters and f_R, f_{WS} and f_{WD} are dummy variables (based on a crude discretisation, see Section 2.2) for respectively rain presence, wind speed levels and wind direction sectors.

Univariate logistic models

In this section, we present results from separate logistic regressions using each available independent variable, together with some possible transformations of these latter.

Models with untransformed variables. The regression curves are presented in Figures 4.2(a)-4.2(f) and the regression parameters are given in Table 4.2. From this we observe statistical significance ($p < 0.001$) of each of the variables tested. The model with θ_{out} has the largest likelihood ratio statistic, implying that it best describes the variations of our outcome variable.

However, as discussed in Section 3.2.3, statistical significance itself does not necessarily provide clear-cut conclusions concerning the model's capacity to correctly explain our outcome variable. We therefore give in Table 4.3 a summary of the possible criteria of goodness-of-fit for each of these models. According to these goodness-of-fit criteria, the model with θ_{out} once again offers the best fit among all variables. We thus conclude that θ_{out} should be integrated in a final model, possibly in conjunction with other variables if their contributions are statistically significant and improve the quality of adjustment. The implications of this superiority of θ_{out} as a predictive variable are discussed in Section 4.3.1.

Models using polynomial logits. We noticed in Figure 4.2(b) that a linear logit does not predict well the observed proportions of windows open at high outdoor temperatures. In order to account for this phenomenon, a possible refinement would be to use a polynomial of degree q for the logit of the probability. In this case:

$$\text{logit}(p) = a + b_1\theta_{\text{out}} + b_2\theta_{\text{out}}^2 + \ldots + b_q\theta_{\text{out}}^q, \tag{4.2}$$

where we use stepwise logistic regression to determine the highest significant order. This procedure determines that a fourth degree polynomial is appropriate, with regression parameters $a = -2.387 \pm 0.005$, $b_1 = (5.55 \pm 0.15) \cdot 10^{-2}$, $b_2 = (1.73 \pm 0.21) \cdot 10^{-3}$, $b_3 = (2.88 \pm 0.11) \cdot 10^{-4}$, $b_4 = (-9.64 \pm 0.19) \cdot 10^{-6}$. We see in Figure 4.2(b) that the associated probability distribution fits better the observed proportions (blue curve); particularly for high values of θ_{out}. Furthermore, regressions with polynomials of lower degree do not offer clear improvements compared to the linear logit model. Although goodness-of-fit indicators are not much improved (Table 4.3), this is the best model when using one sole predictor. Similarly, a limited but significant improvement is obtained when using a polynomial logit with θ_{in} as predictor (Figure 4.2(a)).

Although these models appear to better emulate observed trends, models with polynomial terms tend to be criticised because of the lack of interpretability of their regression coefficients; other approaches are therefore often preferred. Given the structure of the observed proportions, viable possibilities include a non-parametric estimation of the probability, or the fitting of two linear logistic models for the distinct domains of θ_{out} where the observed behaviours are different.

Models based on deviations from comfort temperature. Another possible choice for a driving variable is to use the deviation between θ_{in} or θ_{out} and a comfort temperature $\theta_{\text{in,comf}}$; for example defined by the adaptive comfort model of the CEN or ASHRAE standards (see Section 6.1.1). We perform logistic regression with $(\theta - \theta_{\text{in,comf}})$ as a driving variable, alternatively with $\theta = \theta_{\text{in}}$ and $\theta = \theta_{\text{out}}$. The corresponding results are given in Figure 4.3 and Table 4.2 (bottom).

In this case, we obtain slightly lower goodness of fit and likelihood ratio; that is, the quality of adjustment is somewhat lower than when using raw thermal variables. It is however worth noting that the proportion of windows open reaches a maximum near $\theta_{\text{out}} = \theta_{\text{in,comf}}$. The use of the equations given alternatively by the CEN or ASHRAE standards produce similar results.

Multivariate logistic models

Following from these univariate models, we proceed to consider models with several variables and assess their increased predictive value. In this we determine the best model

Variables	LR	AUC	D_{xy}	Γ	τ_a	R_N^2	B
θ_{out}	330873	0.782	0.563	0.565	0.240	0.273	0.168
θ_{in}	71243	0.632	0.264	0.269	0.113	0.064	0.202
ϕ_{out}	32566	0.590	0.179	0.181	0.076	0.030	0.208
v_{wind}	10992	0.538	0.077	0.078	0.033	0.010	0.212
$p_{atm,red}$	9556	0.559	0.118	0.122	0.050	0.009	0.212
α_{wind}	4153	0.531	0.063	0.065	0.027	0.004	0.213
D_{prec}	202	0.493	-0.013	-0.112	-0.006	0.000	0.213
f_{WD}	27756	0.579	0.157	0.211	0.067	0.025	0.209
f_{WS}	19126	0.566	0.133	0.176	0.057	0.017	0.211
f_R	1065	0.507	0.014	0.119	0.006	0.001	0.213
θ_{out} (polyn.)	349191	0.783	0.566	0.568	0.241	0.287	0.166
θ_{in} (polyn.)	91047	0.637	0.274	0.281	0.117	0.081	0.200
$\theta_{out} - \theta_{comf}^{(CEN)}$	309681	0.774	0.548	0.550	0.234	0.258	0.171
$\theta_{out} - \theta_{comf}^{(ASHRAE)}$	308337	0.774	0.548	0.549	0.234	0.257	0.171
$\theta_{in} - \theta_{comf}^{(CEN)}$	47692	0.603	0.206	0.208	0.088	0.043	0.206
$\theta_{in} - \theta_{comf}^{(ASHRAE)}$	44142	0.602	0.203	0.205	0.087	0.040	0.207
θ_{out}, θ_{in}	343507	0.785	0.570	0.571	0.243	0.283	0.167
θ_{out}, ϕ_{out}	342396	0.785	0.569	0.570	0.243	0.283	0.167
θ_{out}, α_{wind}	334066	0.783	0.566	0.567	0.241	0.276	0.168
θ_{out}, v_{wind}	331803	0.782	0.564	0.566	0.241	0.274	0.168
θ_{out}, D_{prec}	331616	0.782	0.564	0.566	0.240	0.274	0.168
θ_{out}, f_{WD}	332478	0.782	0.565	0.566	0.241	0.275	0.168
θ_{out}, f_{WS}	331683	0.782	0.564	0.565	0.240	0.274	0.168
θ_{out}, f_R	331325	0.782	0.564	0.566	0.240	0.274	0.168
θ_{out}, θ_{in}, ϕ_{out}	354434	0.789	0.577	0.578	0.246	0.291	0.165

Table 4.3: Goodness-of-fit estimators for logistic models including one or several variables

containing two variables, and provided the significance of the added variable and the stability of the primary variable; continuing this procedure for other predictors until no added significance is obtained. This procedure is known as *forward selection* (see Section 3.2.4).

Models with two variables. Based on logistic regression for models including together θ_{out} and each other available variable, we observe (Table 4.3) that the model with θ_{out} and θ_{in} ($a = 0.794 \pm 0.030$, $b_{\text{out}} = 0.14760 \pm 0.00031$, $b_{\text{in}} = -0.1541 \pm 0.0013$) has the highest statistical significance, according to the likelihood ratio statistic. Furthermore, this model is the one that fits best the data, according to all our statistical criteria; but the improvement to these indicators from adding θ_{in} is rather modest. However a plot of the observed proportions of windows open versus θ_{out} and θ_{in}, with regression surface levels (Figure 4.4(a)), shows that observed variations are better accounted for and thus confirms the existence of an independent contribution of each variable. Finally, the stability of the slope associated with θ_{out} is preserved, as its standard error remains extremely low, which shows that the correlation between θ_{in} and θ_{out} is not problematic for this model.

Models with three or more variables. Now that the model including θ_{out} and θ_{in} is retained, we check for the significance of the inclusion of a third parameter. Based on regression results for the models with a third variable, the best model includes the external relative humidity ϕ_{out} and this inclusion is statistically significant ($p < 0.001$). However the goodness-of-fit criteria increase only very slightly (Table 4.3); that is, the added predictive accuracy from the inclusion of ϕ_{out} is marginal. Some other parameters in models with four or five variables were also found to be statistically significant, but without any increase in the goodness-of-fit indicators. For the sake of parsimony, it seems sensible to keep the model with just the two variables θ_{out} and θ_{in}.

Other factors. Inspired by the results of Herkel et al. [81, 82], we attempted to include a factor with twelve levels corresponding to each month of the year, in order to check the existence of an additional effect of season on window actions. This was not found to bring any significant improvement; that is we observe almost the same regression parameters based on θ_{out} for every month.

4.2.3 Model based on a discrete-time Markov process

As noted earlier, a single probability distribution ignores the real dynamic processes leading occupants to perform actions, as the data used to infer them are aggregated observations of window states, but not actual opening or closing actions[7]. In other words these models do not describe an actual probability of opening or closing, but a probability for a window to be "found" open, provided relevant physical parameters. Furthermore it ignores the particular patterns caused by occupancy events, like arrivals or departures of occupants. We thus present in this section an alternative dynamic modelling approach to account for the real adaptive processes of occupants.

Guided by the initial observation that occupancy events have an influence on actions (Section 4.2.1), we may infer different transition probabilities P_{ij} for these events, so that we have three different sub-models for actions on arrival, at departure and during occupancy, as proposed in [76, 82]. Simulation may then be conducted as presented in

[7] The approach used in the Humphreys algorithm [70, 77] is a possible adjustment choice to include dynamics in such probability distributions, although not based on observed actions.

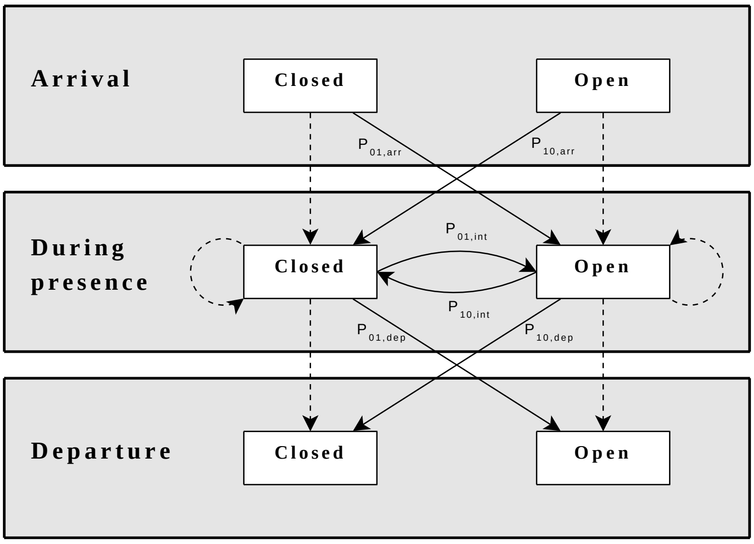


Figure 4.5: General scheme of the Markov process

Figure 4.5: depending on the inital state of the window, opening on arrival is predicted by a specific probability $P_{01,\mathrm{arr}}$, and closing on arrival by $P_{10,\mathrm{arr}}$. Actions after arrival are predicted by another sub-model launched at regular time steps, with transition probabilities $P_{01,\mathrm{int}}$ if the window is closed at this time and $P_{10,\mathrm{int}}$ if it is open. When the occupant leaves his office, a third sub-model predicts actions on departure, with transition probabilities $P_{01,\mathrm{dep}}$ and $P_{10,\mathrm{dep}}$. In each case, P_{00} and P_{11} are easily deduced: $P_{00} = 1 - P_{01}$ and $P_{11} = 1 - P_{10}$.

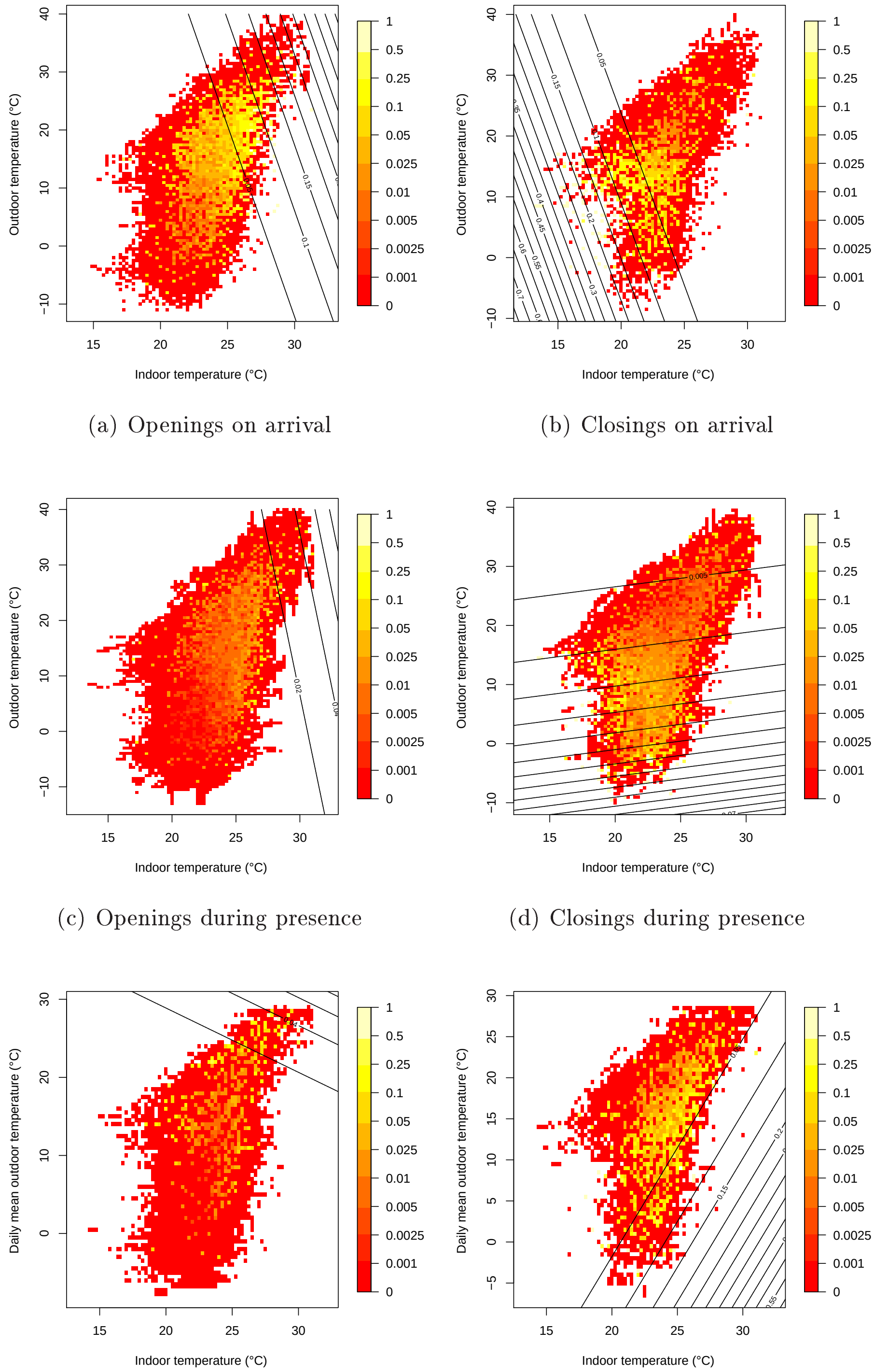
For each sub-model, we filter the data to retain observations related to the relevant occupancy status, and perform logistic regressions on the most relevant environmental parameters; retaining the optimal set of variables in this adapted version of Equation 4.1:

$$\begin{aligned}\mathrm{logit}(p) = a &+ b_{\mathrm{in}}\theta_{\mathrm{in}} + b_{\mathrm{out}}\theta_{\mathrm{out}} + b_{\mathrm{out,dm}}\theta_{\mathrm{out,dm}} + b_R f_R + b_{WS} f_{WS} + b_{WD} f_{WD} \\ &+ b_{\mathrm{GF}} f_{\mathrm{GF}} + b_{\mathrm{pres}} T_{\mathrm{pres}} + b_{\mathrm{abs,prev}} f_{\mathrm{abs,prev}} + b_{\mathrm{abs,next}} f_{\mathrm{abs,next}}, \end{aligned} \tag{4.3}$$

where T_{pres} is the ongoing presence duration, $f_{\mathrm{abs,prev}}$, $f_{\mathrm{abs,next}}$ and f_{GF} are binary variables equal to one respectively for preceding or following absences longer than 8 hours and for offices not on ground floors, and b_{pres}, $b_{\mathrm{abs,prev}}$, $b_{\mathrm{abs,next}}$, b_{GF} are their associated regression parameters.

Sub-model for actions on arrival

Based on our preliminary observations, we will include all actions performed within 5 minutes of arrival in this sub-model. Using the variable selection procedure presented in Section 4.2.2, we conclude that the best model with a single variable uses θ_{in} as a predictor for openings and for closings, while the second most influential physical variable is θ_{out}. We show regression results in Table 4.4 (top) and in Figures 4.6(a)-4.6(b) the observed proportions of actions with contour levels of regression surface. From these results it is clear that θ_{in} exerts the dominant influence on both the opening and closing of windows on arrival.



(a) Openings on arrival

(b) Closings on arrival

(c) Openings during presence

(d) Closings during presence

(e) Openings on departure

(f) Closings on departure

Figure 4.6: Observed transition probabilites (given on a quasi-logarithmic scale), versus bins of indoor and outdoor temperature, with contour lines of equal fitted probabilities

Type	Param.	Opening probabilities			Closing probabilities		
		Estimate	Z	χ^2	Estimate	Z	χ^2
Arrival	a	-13.88 ± 0.37	-37.86		3.97 ± 0.37	10.86	
	b_{in}	0.312 ± 0.016	19.60	384.09	-0.286 ± 0.017	-16.84	283.44
	b_{out}	0.0433 ± 0.0033	13.13	172.38	-0.0505 ± 0.0045	-11.29	127.56
	$b_{\text{abs,prev}}$	1.862 ± 0.044	42.45	1801.91			
	b_R	-0.45 ± 0.11	-3.97	15.79			
Interm.	a	-12.23 ± 0.28	-43.49		-1.64 ± 0.22	-7.57	
	b_{in}	0.281 ± 0.013	22.45	504.03	-0.0481 ± 0.0098	-4.91	24.14
	b_{out}	0.0271 ± 0.0024	11.30	127.74	-0.0779 ± 0.0020	-38.07	1449.26
	b_{pres}	(-8.78 ± 0.53) $\cdot 10^{-4}$	-16.61	275.85	(-1.621 ± 0.059) $\cdot 10^{-3}$	-27.69	766.60
	b_R	-0.336 ± 0.081	-4.13	17.05			
Departure	a	-8.75 ± 0.22	-39.88		-8.54 ± 0.48	-17.83	
	b_{in}				0.213 ± 0.022	9.69	93.95
	$b_{\text{out,dm}}$	0.1371 ± 0.0075	18.65	347.69	-0.0911 ± 0.0061	-14.94	223.19
	$b_{\text{abs,next}}$	0.84 ± 0.12	7.23	52.32	1.614 ± 0.069	23.34	544.86
	b_{GF}	0.83 ± 0.13	6.32	39.96	-0.923 ± 0.068	-13.57	184.18

Table 4.4: Regression parameters, Wald Z and analysis of deviance table for final transition probabilities ($p < 0.001$ for all Wald Z and χ^2 tests)

For $P_{01,\text{arr}}$, the presence of rain is a significant factor, but prior absence duration has a stronger influence. With these variables, Equation 4.3 becomes:

$$\text{logit}\Big(P_{01,\text{arr}}(\theta_{\text{in}}, \theta_{\text{out}}, f_{\text{abs,prev}}, f_R)\Big) = a + b_{\text{in}}\theta_{\text{in}} + b_{\text{out}}\theta_{\text{out}} + b_{\text{abs,prev}} f_{\text{abs,prev}} + b_R f_R. \quad (4.4)$$

On the contrary, we have not found any other significant variable for $P_{10,\text{arr}}$. The final model is thus:

$$\text{logit}\Big(P_{10,\text{arr}}(\theta_{\text{in}}, \theta_{\text{out}})\Big) = a + b_{\text{in}}\theta_{\text{in}} + b_{\text{out}}\theta_{\text{out}}. \quad (4.5)$$

Goodness-of-fit indicators are provided in Table 4.5. There are noticeable differences in behaviour between offices, which will be discussed in Section 4.2.5.

Sub-model for actions during occupancy

We noticed in our preliminary observations that actions during occupancy were extremely rare; particularly openings. We see that θ_{in} is the main driving variable for $P_{01,\text{int}}$, while θ_{out} dominates for $P_{10,\text{int}}$ (indeed θ_{in} is barely significant in this case). This shows that θ_{in} is the real underlying stimulus for openings, while θ_{out} (linked to the feedback of the opening) determines primarily the probability of closing (eg. to prevent over- or underheating), see Figures 4.6(c)-4.6(d).

Rainfall and wind are not significant for openings during occupancy. Both current occupancy duration and the occurrence of rain are found to be significant for $P_{01,\text{int}}$. We thus have the following models,

$$\text{logit}\Big(P_{01,\text{int}}(\theta_{\text{in}}, \theta_{\text{out}}, T_{\text{pres}}, f_R)\Big) = a + b_{\text{in}}\theta_{\text{in}} + b_{\text{out}}\theta_{\text{out}} + b_{\text{pres}} T_{\text{pres}} + b_R f_R, \quad (4.6)$$

$$\text{logit}\Big(P_{10,\text{int}}(\theta_{\text{in}}, \theta_{\text{out}}, T_{\text{pres}})\Big) = a + b_{\text{in}}\theta_{\text{in}} + b_{\text{out}}\theta_{\text{out}} + b_{\text{pres}} T_{\text{pres}}, \quad (4.7)$$

with regression parameters given in Table 4.4 (middle). Goodness of fit indicators are lower than for the sub-model relating to actions on arrival (Table 4.5). One final observation is that transition probabilities remain in both cases very close to zero, and thus consecutive

Model, variables	**LR**	**AUC**	$\mathbf{D_{xy}}$	$\mathbf{\Gamma}$	$\tau_\mathbf{a}$	$\mathbf{R^2_N}$	**B**
$\mathbf{P_{01,arr}}$							
θ_{in}	1183.93	0.699	0.397	0.412	0.019	0.057	0.024
θ_{out}	1051.78	0.702	0.403	0.418	0.019	0.050	0.024
θ_{in}, θ_{out}	1352.38	0.724	0.447	0.462	0.021	0.065	0.024
θ_{in}, θ_{out}, f_R	1369.29	0.725	0.449	0.464	0.021	0.065	0.024
θ_{in}, θ_{out}, f_{abs}, f_R	2916.73	0.782	0.563	0.578	0.027	0.138	0.023
$\mathbf{P_{10,arr}}$							
θ_{in}	531.61	0.687	0.374	0.385	0.022	0.066	0.028
θ_{out}	388.35	0.708	0.415	0.429	0.024	0.048	0.029
θ_{in}, θ_{out}	651.34	0.718	0.437	0.446	0.025	0.080	0.028
$\mathbf{P_{01,int}}$							
θ_{in}	1455.01	0.666	0.331	0.435	0.003	0.030	0.004
θ_{out}	1063.67	0.640	0.280	0.383	0.002	0.022	0.004
θ_{in}, θ_{out}	1559.35	0.677	0.353	0.460	0.003	0.032	0.004
θ_{in}, θ_{out}, f_R	1579.31	0.679	0.358	0.464	0.003	0.032	0.004
θ_{in}, θ_{out}, T_{pres}, f_R	2135.68	0.705	0.410	0.505	0.003	0.044	0.004
$\mathbf{P_{10,int}}$							
θ_{in}	428.27	0.592	0.184	0.224	0.004	0.009	0.011
θ_{out}	1336.61	0.667	0.334	0.372	0.007	0.027	0.011
θ_{in}, θ_{out}	1340.26	0.665	0.331	0.369	0.007	0.027	0.011
θ_{in}, θ_{out}, T_{pres}	4134.70	0.764	0.528	0.555	0.011	0.083	0.011
$\mathbf{P_{01,dep}}$							
$\theta_{\text{out,dm}}$	372.92	0.736	0.472	0.543	0.004	0.07	0.004
θ_{in}, $\theta_{\text{out,dm}}$	382.80	0.746	0.492	0.565	0.004	0.072	0.004
θ_{in}, $\theta_{\text{out,dm}}$, f_{abs}, f_{GF}	464.64	0.749	0.498	0.572	0.004	0.087	0.004
$\theta_{\text{out,dm}}$, f_{abs}, f_{GF}	462.17	0.741	0.482	0.557	0.004	0.086	0.004
$\mathbf{P_{10,dep}}$							
$\theta_{\text{out,dm}}$	93.49	0.607	0.213	0.230	0.012	0.012	0.028
θ_{in}, $\theta_{\text{out,dm}}$	206.84	0.648	0.296	0.310	0.017	0.026	0.028
θ_{in}, $\theta_{\text{out,dm}}$, f_{abs}	640.04	0.719	0.438	0.457	0.025	0.079	0.027
θ_{in}, $\theta_{\text{out,dm}}$, f_{abs}, f_{GF}	813.61	0.736	0.472	0.486	0.027	0.100	0.027

Table 4.5: Goodness-of-fit estimators for Markovian transition probabilities

repeated predictions of the same state are very likely so that the associated computation is wasteful. An alternative would be to increase the time step, but this would result in neglecting openings of short duration or artificially increasing the duration of other openings. A more appropriate method for intermediate actions is proposed in Section 4.2.4.

Sub-model for actions at departure

Actions on departure are of a different nature: their goal is not to modify the indoor environment in order to favour thermal comfort for immediate further occupancy. They may therefore be influenced by other factors, such as the predicted duration of subsequent absence, the desire to induce night ventilation or by security issues. We thus include the observed subsequent absence duration in our model. For both openings and closings, we find that the daily mean outdoor temperature $\theta_{\text{out,dm}}$ fits better the data than θ_{out}, and that the position within building (at or above the ground floor) is a significant parameter. In the case of openings on departure, θ_{in} is not a significant parameter. We retain thus the following transition probabilities,

$$\begin{aligned} &\operatorname{logit}\Big(P_{01,\text{dep}}(\theta_{\text{out,dm}}, f_{\text{abs,next}}, f_{GF})\Big) \\ &= a + b_{\text{out,dm}}\theta_{\text{out,dm}} + b_{\text{abs,next}} f_{\text{abs,next}} + b_{GF} f_{GF}, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \operatorname{logit}\Big(P_{10,\text{dep}}(\theta_{\text{in}}, \theta_{\text{out,dm}}, f_{\text{abs,next}}, f_{GF})\Big) &= a + b_{\text{in}}\theta_{\text{in}} \\ &+ b_{\text{out,dm}}\theta_{\text{out,dm}} + b_{\text{abs,next}} f_{\text{abs,next}} + b_{GF} f_{GF}, \end{aligned} \tag{4.9}$$

with regression parameters given in Table 4.4 (bottom). Analysis of deviance shows that the duration of absence and the daily mean outdoor temperature are the most influential factors for openings and for closings.

Remarks

By defining factors for previous and next absence durations, our final sub-models for actions on arrival and at departure may be seen in fact as divided into separate models for short and long absences. This implies that the start of a long absence period increases the probability of closing at departure, which is an expected result. The observation that a long absence preceding an arrival increases the probability of opening could be explained by the fact that, in this case, odours might have accumulated in the office, which could be partly a consequence that most occupants close their doors when leaving for a long absence. Finally, we have confirmed that thermal stimuli are the key variables influencing actual actions on windows.

Our sub-model for actions on departure takes the following absence duration as input, and thus the future value of a stochastic variable. This implies that a model of occupancy presence should be a pre-process (for the entire simulation period) to the model of window opening. This is acceptable since occupancy is clearly independent from window state.

Goodness-of-fit criteria show that our sub-models do not offer equal performances (Table 4.5), this being lowest for actions during presence.

1^st^ state	**2^nd^ state**	**Prop. for 3^rd^ state**	
		Closed	Open
Closed	**Closed**	0.933	0.067
	Open	0.555	0.445
Open	**Closed**	0.555	0.445
	Open	0.131	0.869

Table 4.6: Observed transition probabilities for second order Markov models

The limit model

The retained model corresponds thus to a classical non-stationary two-state Markov chain, which can be written in matrix notation (Section 3.3.2):

$$P = \begin{pmatrix} 1 - P_{01} & P_{01} \\ P_{10} & 1 - P_{10} \end{pmatrix}, \quad \text{with } 0 < P_{01}, P_{10} < 1. \tag{4.10}$$

It may be shown that as $n \to \infty$, the n-step transition probability becomes

$$P^n \to \frac{1}{P_{01} + P_{10}} \begin{pmatrix} P_{10} & P_{01} \\ P_{10} & P_{01} \end{pmatrix} = \mathbf{I} \frac{1}{P_{01} + P_{10}} \begin{pmatrix} P_{10} \\ P_{01} \end{pmatrix}, \tag{4.11}$$

and thus the overall probability for a window to be open is given by $P_{01}/(P_{01} + P_{10})$. However, as our model is based on occupancy-dependent transition probabilities we are not able to explicitly compute the limit model in this case. Nevertheless, the limit $P_{01}/(P_{01} + P_{10})$ may be interpreted as the static probability distributions derived in Section 4.2.2. Therefore this first approach is simply a particular case of the Markov model, which itself provides for a higher degree of modelling detail.

Higher-order Markov models

We assess now the significance of higher-order Markov chains (see Section 3.3.2) in this situation. Neglecting the influence of driving variables, a zeroth-order chain is based on the overall observed proportion of open windows, that is $P_0 = 0.308$, $P_1 = 1 - p_0 = 0.692$ (Table 2.5, bottom line).

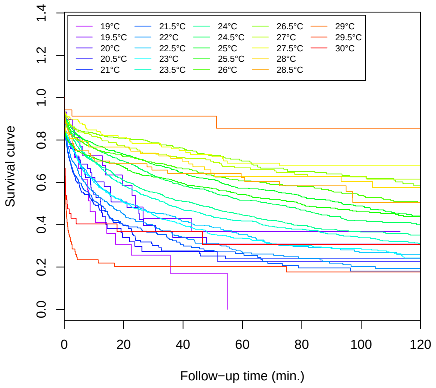
Based on the likelihood ratio statistic deduced from the observed transitions between three consecutive states (Table 4.6), we observe that the second-order transitions P_{ijk} are significant compared to first-order probabilities P_{ij} (the significance of P_{ij} over the zeroth-order model P_i was clear from Table 4.1). This means that the probability P_{0jk} significantly differs from P_{1jk}.

We do not attempt to fit second-order transition probabilities but we rather propose a modelling approach based on continuous-time random processes in Section 4.2.4, which enables us to drop the Markov condition if necessary.

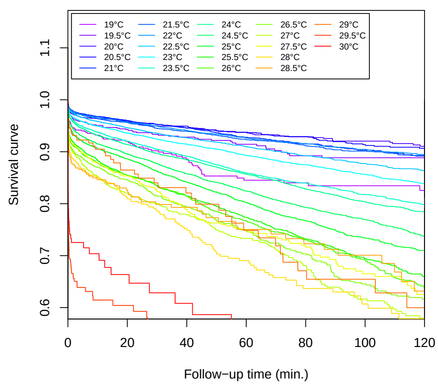
4.2.4 Continuous-time random process

Opening duration

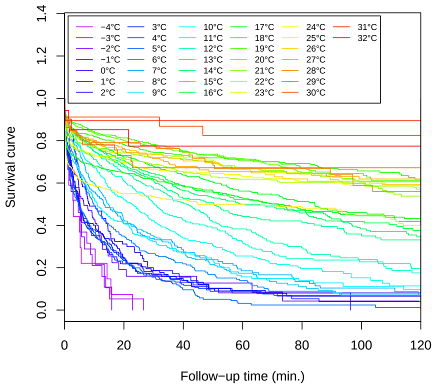
According to the concepts described in Section 3.4, we infer a distribution for the duration during which people leave their window closed following their arrival, and during which the



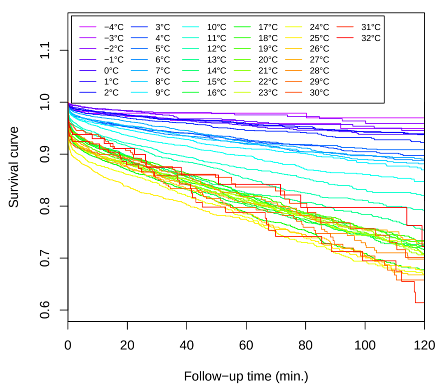
(a) Opening duration: curves based on domains of θ_{in}



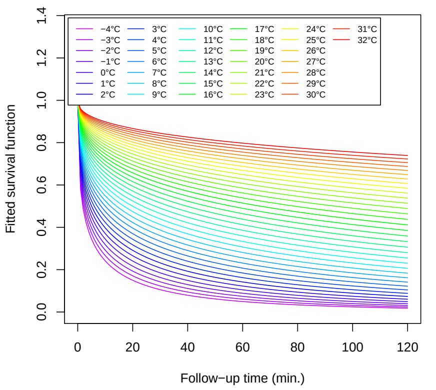
(b) Closing duration: curves based on domains of θ_{in}



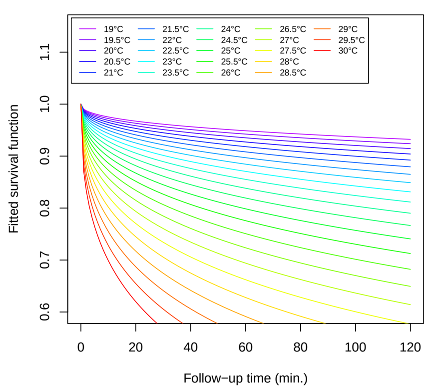
(c) Opening duration: curves based on domains of θ_{out}



(d) Closing duration: curves based on domains of θ_{out}



(e) Opening duration: fit based on θ_{out}



(f) Closing duration: fit based on θ_{in}

Figure 4.7: Kaplan-Meier estimators of survival functions of opening (left) and closing (right) duration. Fitted parametric survival functions are shown in the bottom charts

window is left open. Kaplan-Meier estimates of survival curves are shown in Figures 4.7(a) and 4.7(c), in which each curve refers to an interval[8] of observed initial values of θ_{in} or θ_{out}.

The duration of window openings which were interrupted upon departure needs special treatment. In this case, the reason for closing (or for leaving open) windows is not linked to discomfort, meaning that if the occupant had stayed longer it is not clear that he would have closed the window. We thus have incomplete information: we know that the occupant wished to leave the window open *at least* until this moment. Such opening durations are classified as censored data.

A trend of diminished rates of decay of opening times may be noticed in Figure 4.7(a) when θ_{in} rises until 26°C, while they remain similar above. These decay rates are more clearly differentiated in Figure 4.7(c), which implies that opening durations are more strongly associated with θ_{out}. Both variables are significant ($p < 0.001$) according to the log-rank test.

Detailed analysis of the distribution of opening times shows that the hazard rate $h(t)$ is clearly non-constant and decreases with t, meaning that closings have an increased risk of occurring shortly after openings. Based on the 13489 observed openings, of which 7451 (55.2%) were censored, and using a Weibull distribution (see Section 3.4) we find that the best model with a single variable uses θ_{out} as its predictor ($p < 0.001$, $R^2 = 0.110$). The estimate for the shape is $\log(1/\alpha) = 0.872 \pm 0.011$, while the scale is $\lambda = 1/\exp\big((2.151 \pm 0.066) + (0.1720 \pm 0.0044) \cdot \theta_{\text{out}})\big)$. The fitted survival function is shown in Figure 4.7(e). The variable θ_{in}, if included with this model, is not statistically significant ($p > 0.1$), likewise other potential variables. These results are consistent with our sub-model for window closings during occupancy developed in Section 4.2.3, where θ_{out} is the main driving variable in $P_{10,\text{int}}$.

Closing duration

The data of closing duration include two types of intervals: delay until opening following occupants' arrival, and delay until opening following a prior closing. We see that θ_{out} has less influence than θ_{in} on closing duration (Figure 4.7(b) and 4.7(d)) and therefore on the decay of survival curves, which differ less in the range of values of θ_{out} (Figure 4.7(d)). Conversely, the survival curves vary clearly for different values of θ_{in}. Furthermore, we can straightforwardly interpret the immediate decays along the ordinates in closing durations as *opening probabilities on arrival*, that increase strongly with θ_{in} as expected. Intermediate openings are then described by the rest of the curve, with higher proportional decays being observed for higher temperature.

As for openings, we once again use the Weibull distribution to describe closing durations. We include first θ_{in}, and notice that the addition of θ_{out} is significant. Based on 203881 time intervals with closed windows of which 185277 (90.9%) were censored, we obtain for the shape $\log(1/\alpha) = 0.874 \pm 0.006$, while the scale is $\lambda = 1/\exp\big((16.26 \pm 0.27) + (-0.264 \pm 0.012) \cdot \theta_{\text{in}} + (-0.110 \pm 0.003) \cdot \theta_{\text{out}}\big)$, with $R^2 = 0.030$. For illustration purposes, we show the fitted survival function with θ_{in} as the sole driving variable in Figure 4.7(f).

[8] The intervals of θ_{in} (°C) are set as $[18.75, 19.25)$, $[19.25, 19.75)$, …, $[29.75, 30.25)$, and for θ_{out} (°C), $[-4.5, 3.5)$, $[-3.5, -2.5)$, …, $[31.5, 32.5)$. These intervals of θ_{in} (resp. θ_{out}) cover 99.5% (resp. 99.7%) of openings.

Remarks

We observe that opening and closing durations are poorly fitted by an exponential distribution. The memoryless property referenced above that defines discrete-time Markov processes is thus not fulfilled, which suggests that in this context window opening and closing processes are not memoryless, and so are not fully appropriately modelled by a Markov process.

The obtained Weibull distributions confirm once again that delayed opening of windows is mainly caused by indoor stimuli, while the main driving stimulus for window closings is external (the feedback of the opening).

The main limitation in the modelling of actions with survival curves is the obvious risk in predicting in advance potentially long opening times, independently of subsequent variations of environmental stimuli. For instance, as a transition from closed to open is performed, the indoor conditions evolve in response to heat transfers with the outdoors. A sensible compromise could be to model survival time up to a reasonable horizon, and then reset the survival distribution to shift to the curve adjusted to the new environmental conditions. However, this procedure would be fully rigorous only if the memoryless property $P(T > s + t | T > s) = P(T > t)$ was verified; only the exponential distribution satisfies this condition. However, as 90% of occupancy durations are shorter than 105 minutes, the prediction of problematically long opening durations will occur very rarely, so that this compromise would appear not to be necessary.

4.2.5 Integration of individual behaviours

If everyone is thinking alike, then somebody isn't thinking.
Attributed to Gen. George S. Patton (1885-1945)

The models developed above were derived from data relating to the whole set of occupants and for the entire surveyed period. We examine here variations in behaviour among the surveyed occupants and provide a method to account for the observed behavioural diversity.

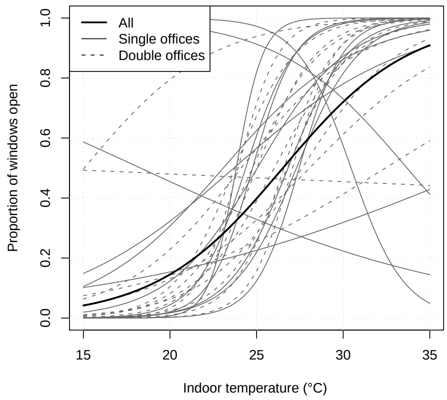
Variability between occupants

We provide in Table 4.7 the principal indicators concerning general conditions and behavioural differences for all the surveyed occupants (or their combinations). In order to distinguish between "active" and "passive" occupants a possibility is to use as an indicator the percentage of time occupants leave their windows open during occupancy. We see that this ratio ranges from 13.5% to 47.7% with respect to sufficiently long occupancy durations (see Table 2.5), with half of the occupants grouped between 19.5% and 42.7%.

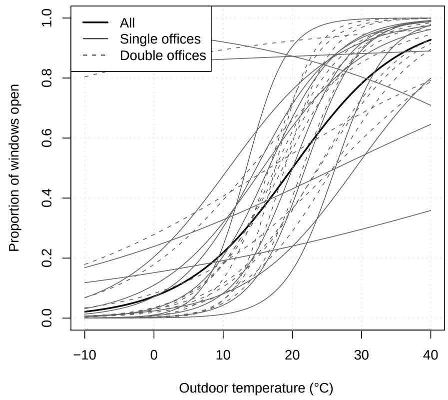
Furthermore, different occupants may vary not only in the intensity of their behaviour but also in its nature. For example in Figure 4.8 we present the obtained logistic models for different occupants. It can be seen that a minority of occupants is weakly influenced by thermal stimuli (the probability to observe their window to be open varies little with θ_{out} and θ_{in}). Nevertheless, for the majority of them, the slopes of the univariate models for these variables are not very different; the differences arising mainly in the intercepts. Variability may thus be meaningfully summarised, as a first approximation, by a single parameter: the characteristic temperature $\theta_{50} = -a/b$.

Ref.	**Nb. Pers.**	**Survey duration**	**Ratio open**	**Regr. param.** $\mathbf{a_{out}}$	$\mathbf{b_{out}}$	$\boldsymbol{\theta_{out,50}}$ **[°C]**	$\boldsymbol{\theta_{in,50}}$ **[°C]**	**Predictive**
001-04-27	2	730	0.210	-3.721	0.157	23.7	26.8	No
001-28-X	2	699	0.343	-3.532	0.202	17.5	25.0	No
001-17-30	2	931	0.459	-1.522	0.109	14.0	24.1	No
001-17-35	2	206	0.460	-3.304	0.178	18.6	24.8	Yes
002-21-24	2	275	0.400	-6.431	0.367	17.5	23.7	No
002-16-25	2	334	0.470	-0.951	0.057	16.6	12.3	Yes
002-01-16	2	153	0.051	-5.506	0.256	21.5	27.1	No
002-11-16	2	496	0.178	-3.941	0.152	25.9	26.6	Yes
003-07	1	2808	0.159	-3.703	0.127	29.1	26.8	No
004-36	2	321	0.181	-5.818	0.308	18.9	32.4	No
004-26-36	1	2292	0.263	-4.804	0.240	20.0	24.4	No
101-15-28	2	606	0.188	-3.497	0.147	23.8	27.6	No
101-22	1	1896	0.172	-5.703	0.261	21.9	27.3	No
103-23	1	272	0.135	-7.011	0.268	26.2	27.5	No
103-03	1	2191	0.348	-1.159	0.044	26.4	18.3	No
103-42*	1	195	0.913	2.963	-0.052	NA	NA	Yes
103-29*	1	56	0.874	1.750	0.009	NA	NA	Yes
104-19*	1	2808	0.204	-1.724	0.029	NA	NA	Yes
105-38	1	2808	0.380	-2.091	0.133	15.7	25.1	No
106-06-39	2	2808	0.326	-3.362	0.187	18.0	26.0	No
201-31	1	2808	0.427	-2.573	0.174	14.8	23.1	Yes
202-05-23	2	457	0.195	-4.751	0.286	16.6	23.9	No
203-09	2	1550	0.246	-2.411	0.093	25.9	28.1	Yes
203-08-14	1	517	0.249	-4.537	0.340	13.4	23.8	No
203-12-40*	2	147	0.930	1.773	0.036	NA	NA	Yes
204-18	1	1048	0.381	-3.392	0.205	16.6	24.8	No
204-10	1	1759	0.477	-1.360	0.126	10.8	23.7	No

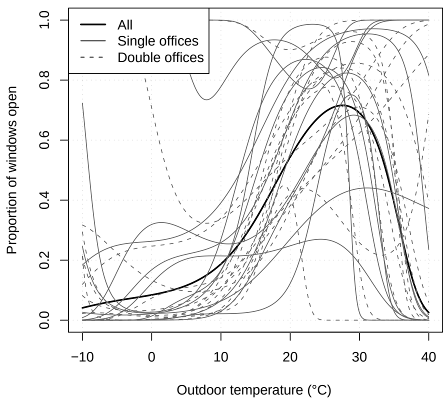
Table 4.7: Variability between occupants: reference, number of occupants, duration of observations (days), overall proportion of time open, logistic regression parameters using θ_{out}, indoor and outdoor characteristic temperatures and presence of predictive behaviour. Asterisks show occupants for which at least one regression parameter was not statistically significant.



(a) Linear logistic models based on indoor temperature

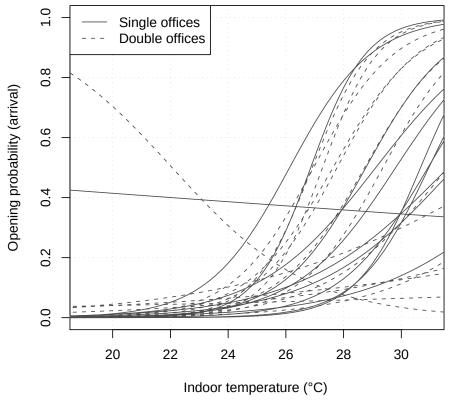


(b) Linear logistic models based on outdoor temperature

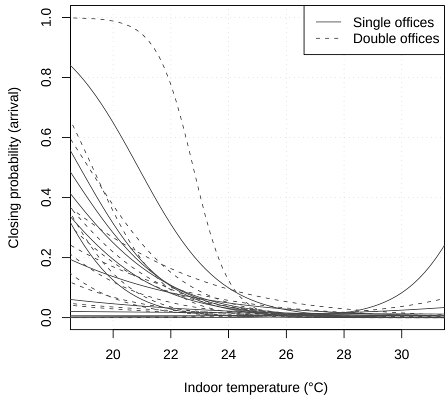


(c) Polynomial logistic models based on outdoor temperature

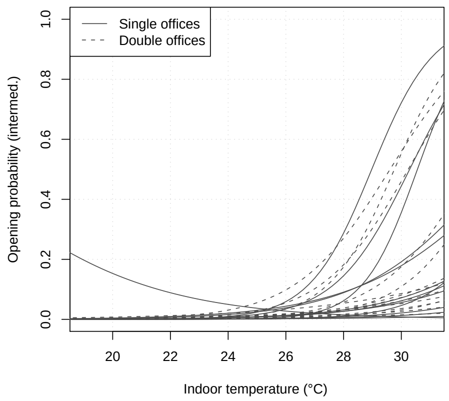
Figure 4.8: Occupant specific logistic models



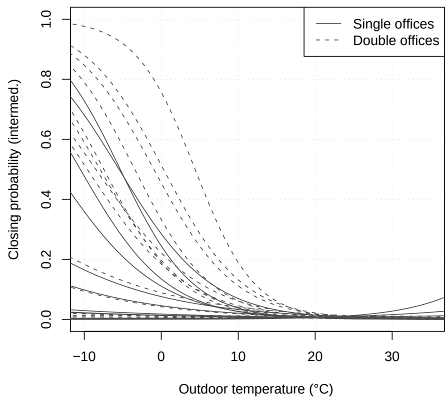
(a) Opening on arrival



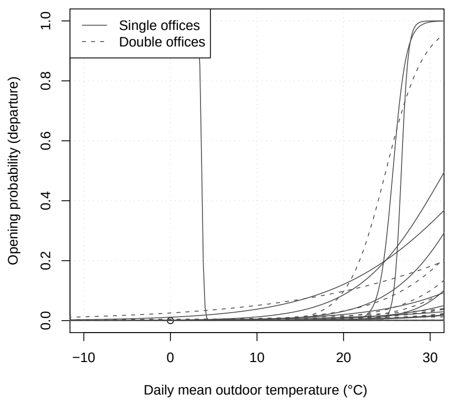
(b) Closing on arrival



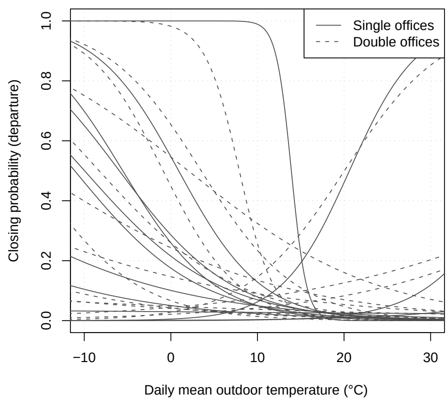
(c) Opening during occupancy



(d) Closing during occupancy



(e) Opening on departure



(f) Closing on departure

Figure 4.9: Occupant specific action probabilities based on the most influential physical variable

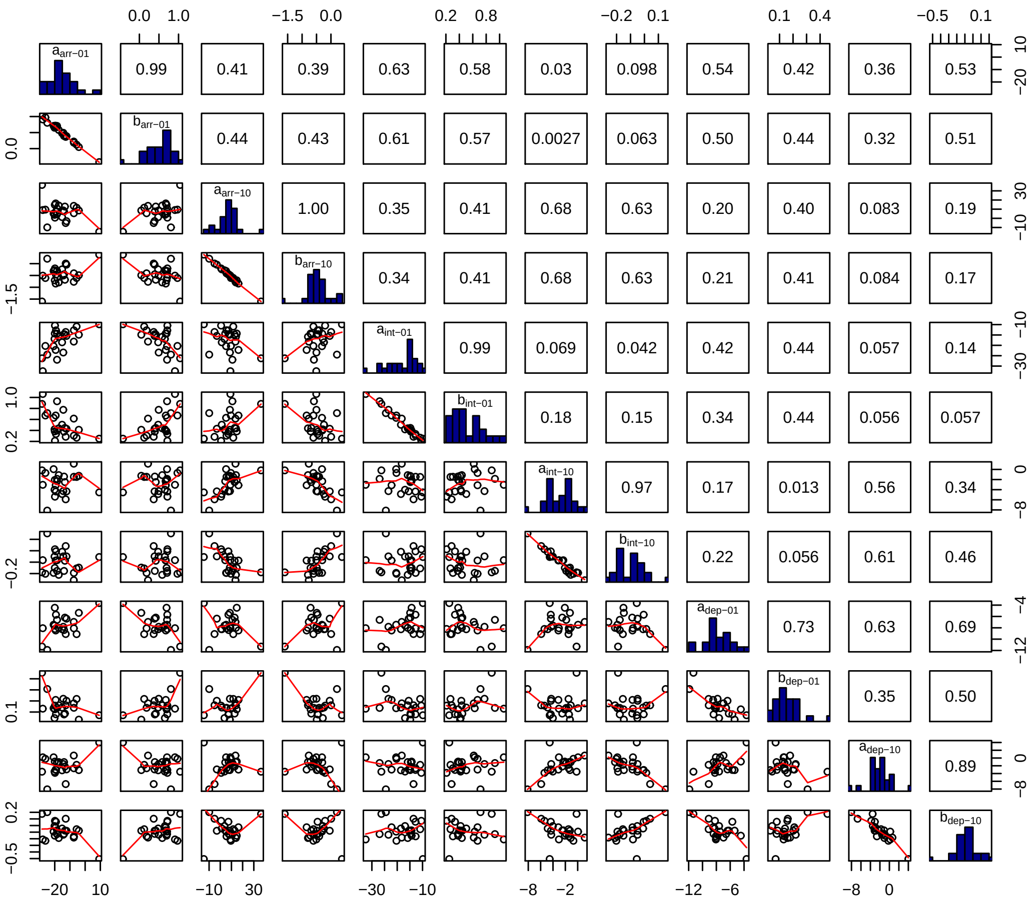


Figure 4.10: Bivariate plots between all individual regression parameters, with local polynomial regression, correlations and histograms

It is worth noticing that some occupants have open windows at high θ_{out}, while others follow the decreasing trend shown in Figure 4.2(b) to prevent the incoming of hot outside air. We qualify this behaviour as "predictive" if the polynomial terms in the logit are significant (see Table 4.7). We see that nine observed occupants adopt this preventive strategy.

Similarly, differences in behavioural patterns between occupants may be noticed if we derive individual transition probabilities for the discrete-time Markov process based on the most influential physical predictor only (θ_{in} for $P_{01,\text{arr}}$, $P_{10,\text{arr}}$ and $P_{01,\text{int}}$, θ_{out} for $P_{10,\text{int}}$ and $\theta_{\text{out,dm}}$ for $P_{01,\text{dep}}$ and $P_{10,\text{dep}}$). The regression parameters are presented in Table 4.8 and the obtained action probability curves in Figure 4.9. These results suggest that occupants generally display the same type of behaviour, but at higher or lower temperatures. Furthermore, the predictive scheme described above, i.e. refraining from opening windows for hot outside conditions, is reproduced by the same occupants.

Figure 4.10 shows the main patterns linking these regression parameters, which may inform an explicit simulation of individual diversity for integration in simulations. Further

Pers Ref.	$\mathbf{P_{01,arr}}(\theta_{\mathbf{in}})$ a	b	$\mathbf{P_{10,arr}}(\theta_{\mathbf{in}})$ a	b	$\mathbf{P_{01,int}}(\theta_{\mathbf{in}})$ a	b
001-04-27	-13.9 ± 3.9	0.39 ± 0.16	7.5 ± 2.5	-0.45 ± 0.11	-14.6 ± 1.7	0.38 ± 0.07
001-28-X	-19.2 ± 1.6	0.71 ± 0.06	7.3 ± 3.3	-0.46 ± 0.14	-12.8 ± 1.0	0.34 ± 0.04
001-17-30*	-18.5 ± 2.7	0.67 ± 0.11	4.4 ± 4.6	-0.30 ± 0.19	-14.4 ± 2.4	0.40 ± 0.10
001-17-35	-5.6 ± 1.3	0.12 ± 0.06	11.6 ± 1.9	-0.61 ± 0.09	-10.9 ± 1.1	0.27 ± 0.05
002-21-24	-28.4 ± 7.2	1.04 ± 0.29	36.5 ± 19.9	-1.60 ± 0.86	-26.3 ± 3.6	0.88 ± 0.15
002-16-25	-20.7 ± 3.4	0.71 ± 0.13	5.2 ± 6.3	-0.31 ± 0.26	-20.4 ± 1.8	0.63 ± 0.07
002-01-16	-19.9 ± 3.1	0.71 ± 0.12	0.7 ± 4.6	-0.20 ± 0.20	-10.5 ± 2.0	0.21 ± 0.08
002-11-16	-7.5 ± 2.4	0.22 ± 0.10	13.6 ± 2.7	-0.77 ± 0.13	-18.2 ± 1.8	0.61 ± 0.08
003-07	-20.1 ± 1.7	0.65 ± 0.07	3.8 ± 2.4	-0.28 ± 0.10	-16.9 ± 2.0	0.42 ± 0.08
004-36	-17.4 ± 1.4	0.58 ± 0.06	14.0 ± 1.4	-0.80 ± 0.07	-15.9 ± 1.2	0.44 ± 0.05
004-26-36*	-26.2 ± 5.9	0.97 ± 0.26	9.5 ± 5.3	-0.61 ± 0.25	-20.3 ± 5.1	0.67 ± 0.22
101-15-28	-4.1 ± 2.0	0.05 ± 0.09	8.6 ± 2.1	-0.49 ± 0.10	-14.9 ± 1.7	0.41 ± 0.07
101-22	-27.9 ± 2.8	0.91 ± 0.11	9.1 ± 1.7	-0.51 ± 0.07	-32.5 ± 1.9	1.06 ± 0.07
103-23	-12.8 ± 1.3	0.37 ± 0.05	-5.1 ± 4.3	0.01 ± 0.18	-13.6 ± 1.7	0.28 ± 0.07
103-03*	NA	NA	NA	NA	-17.0 ± 6.5	0.44 ± 0.26
103-42*	0.2 ± 9.1	-0.03 ± 0.33	-27.9 ± 14.1	0.85 ± 0.49	4.5 ± 8.6	-0.31 ± 0.31
103-29*	-28.0 ± 9.2	1.04 ± 0.34	0.8 ± 7.0	-0.19 ± 0.30	-13.2 ± 6.8	0.39 ± 0.26
104-19	-24.9 ± 2.1	0.81 ± 0.08	-9.9 ± 4.7	0.21 ± 0.19	-24.5 ± 1.6	0.72 ± 0.06
105-38	-13.3 ± 1.3	0.42 ± 0.05	12.8 ± 1.4	-0.68 ± 0.06	-14.8 ± 1.0	0.44 ± 0.04
106-06-39	-19.4 ± 1.2	0.67 ± 0.05	16.3 ± 1.7	-0.84 ± 0.07	-22.2 ± 1.3	0.67 ± 0.05
201-31	-12.3 ± 1.0	0.39 ± 0.05	-3.1 ± 1.4	-0.04 ± 0.06	-15.5 ± 0.7	0.43 ± 0.03
202-05-33	-7.4 ± 2.7	0.18 ± 0.12	5.6 ± 4.2	-0.41 ± 0.20	-12.7 ± 3.1	0.30 ± 0.14
203-09	-18.5 ± 2.7	0.71 ± 0.12	10.9 ± 4.8	-0.59 ± 0.21	-27.0 ± 3.1	0.93 ± 0.13
203-08-14	-15.2 ± 1.6	0.48 ± 0.07	1.3 ± 2.7	-0.24 ± 0.12	-18.5 ± 1.4	0.50 ± 0.06
203-12-40	9.3 ± 8.4	-0.42 ± 0.35	-14.7 ± 11.7	0.38 ± 0.46	-9.9 ± 4.6	0.25 ± 0.19
204-18	-14.4 ± 1.5	0.49 ± 0.06	8.6 ± 2.1	-0.50 ± 0.09	-23.5 ± 1.7	0.77 ± 0.07
204-10	-20.0 ± 1.9	0.69 ± 0.08	14.7 ± 2.0	-0.71 ± 0.08	-14.3 ± 1.3	0.38 ± 0.05
Pers Ref.	$\mathbf{P_{10,int}}(\theta_{\mathbf{out}})$ a	b	$\mathbf{P_{01,dep}}(\theta_{\mathbf{out,dm}})$ a	b	$\mathbf{P_{10,dep}}(\theta_{\mathbf{out,dm}})$ a	b
001-04-27	-1.4 ± 0.3	-0.18 ± 0.02	-8.4 ± 2.3	0.08 ± 0.15	-1.0 ± 0.7	-0.13 ± 0.05
001-28-X	-4.4 ± 0.1	-0.01 ± 0.01	-5.7 ± 0.7	0.08 ± 0.05	-3.1 ± 0.3	0.06 ± 0.02
001-17-30	-2.3 ± 0.5	-0.08 ± 0.02	NA	NA	-1.1 ± 1.4	-0.07 ± 0.07
001-17-35	-1.5 ± 0.1	-0.15 ± 0.01	-9.1 ± 1.9	0.22 ± 0.11	-1.8 ± 0.4	-0.06 ± 0.03
002-21-24	-0.2 ± 0.4	-0.19 ± 0.04	-11.3 ± 5.5	0.45 ± 0.36	-3.5 ± 2.3	0.18 ± 0.16
002-16-25	0.1 ± 0.4	-0.19 ± 0.02	-7.5 ± 1.5	0.13 ± 0.08	0.2 ± 1.1	-0.09 ± 0.06
002-01-16	-5.4 ± 0.3	0.01 ± 0.02	-8.2 ± 2.8	0.16 ± 0.17	-3.8 ± 0.6	0.07 ± 0.03
002-11-16	1.1 ± 0.3	-0.26 ± 0.02	-7.0 ± 1.3	0.18 ± 0.08	0.6 ± 0.8	-0.18 ± 0.05
003-07	-4.0 ± 0.3	-0.05 ± 0.01	-7.6 ± 0.8	0.05 ± 0.05	-3.5 ± 0.5	-0.01 ± 0.03
004-36	-1.1 ± 0.2	-0.21 ± 0.01	-8.0 ± 0.6	0.16 ± 0.04	-1.0 ± 0.4	-0.19 ± 0.03
004-26-36	-0.7 ± 0.4	-0.20 ± 0.03	NA	NA	-0.2 ± 0.7	-0.23 ± 0.05
101-15-28	-1.3 ± 0.2	-0.15 ± 0.01	-5.4 ± 0.7	0.03 ± 0.05	-3.1 ± 1.0	-0.08 ± 0.07
101-22	-3.0 ± 0.4	-0.08 ± 0.02	-8.1 ± 0.9	0.13 ± 0.06	0.2 ± 0.7	-0.21 ± 0.04
103-23	-5.9 ± 0.2	0.04 ± 0.01	-9.1 ± 1.4	0.16 ± 0.07	-6.7 ± 1.0	0.10 ± 0.05
103-03*	-0.9 ± 1.0	-0.17 ± 0.05	NA	NA	NA	NA
103-42*	-5.5 ± 1.2	0.01 ± 0.05	NA	NA	NA	NA
103-29*	-6.6 ± 0.7	0.08 ± 0.03	NA	NA	-5.3 ± 0.9	0.25 ± 0.05
104-19	-8.1 ± 0.4	0.15 ± 0.02	-11.9 ± 1.5	0.31 ± 0.07	-8.1 ± 1.1	0.20 ± 0.05
105-38	-2.5 ± 0.1	-0.09 ± 0.01	-6.2 ± 0.4	0.09 ± 0.03	-2.2 ± 0.3	-0.08 ± 0.02
106-06-39	-3.1 ± 0.1	-0.09 ± 0.01	-8.5 ± 0.7	0.21 ± 0.04	-3.0 ± 0.3	-0.03 ± 0.02
201-31	-4.3 ± 0.2	-0.05 ± 0.01	-6.4 ± 0.3	0.20 ± 0.01	-2.9 ± 0.3	-0.08 ± 0.02
202-05-33	-1.4 ± 0.2	-0.19 ± 0.02	-8.1 ± 2.0	0.14 ± 0.19	-2.6 ± 1.2	-0.16 ± 0.12
203-09	-1.9 ± 0.2	-0.18 ± 0.02	-6.6 ± 1.1	0.14 ± 0.12	-1.3 ± 0.6	-0.13 ± 0.05
203-08-14	-4.9 ± 0.2	-0.01 ± 0.01	-8.3 ± 1.2	0.14 ± 0.07	-3.2 ± 0.3	-0.04 ± 0.02
203-12-40	-4.5 ± 1.3	-0.06 ± 0.06	-3.6 ± 3.2	0.07 ± 0.16	4.0 ± 5.5	-0.51 ± 0.33
204-18	-4.3 ± 0.1	-0.05 ± 0.01	-7.8 ± 0.8	0.22 ± 0.06	-1.5 ± 0.3	-0.14 ± 0.02
204-10	-2.1 ± 0.2	-0.15 ± 0.01	-4.5 ± 0.2	0.12 ± 0.02	-0.9 ± 0.3	-0.15 ± 0.02

Table 4.8: Occupant specific parameters for action probabilities on windows including a single variable. Asterisks are used to identify occupants for which at least one regression parameter was not statistically significant.

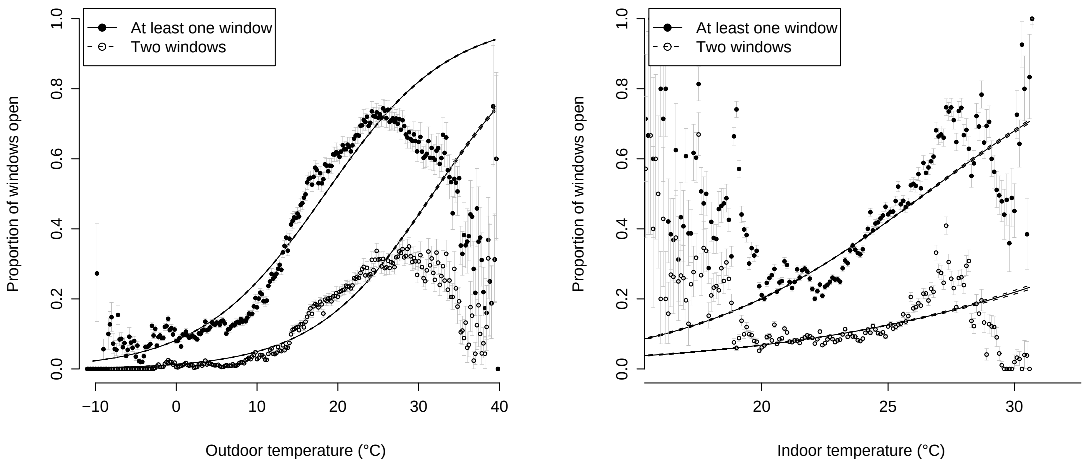


Figure 4.11: Observed proportion of windows open: extension to the case of two windows

discussion is also provided in Section 7.3.2.

Group actions

No clear difference in behaviour related to total opening duration or regression parameters is distinguishable in Tables 4.7 and 4.8 between offices with one or two occupants. In this latter case, one possibility is to assume that occupants act independently and the "activity" in an office is aligned to that of the most active (or assertive) of the occupants. Similarly, none of the distributions displayed in Figures 4.8 and 4.9 show differentiated behaviour between single and double-occupied offices.

When using the discrete or continuous time random processes to predict group actions, two variants may be suggested. A first possibility is to explicitly model the occupancy of each potential occupant, and launch the window opening model for each of them. An alternative would be to adapt the sub-model for actions during presence to the special case where a second arrival may occur in an already occupied office. This sub-model will then predict higher transition probabilities.

4.2.6 Use of several windows

In the individual offices, each occupant has the possibility to freely interact with two windows. In Section 4.2.2 we considered the probability that *at least one window is open*. We may now perform a similar logistic regression – on the subset of individual offices – for the probability that both windows are open. For the binary outcome "at least one window open", the distribution for these offices is $\mathrm{logit}(p) = -2.259 + 0.1172 \cdot \theta_{\mathrm{out}}$, which gives $\theta_{50} = 19.3°\mathrm{C}$, while for two windows open this becomes $\mathrm{logit}(p) = -4.181+0.1288 \cdot \theta_{\mathrm{out}}$, and thus $\theta_{50} = 32.5°\mathrm{C}$. Both distributions and observed proportions are shown in Figure 4.11 in which the slopes are similar. Furthermore, inspection of individual results suggests that this phenomenon may be observed office by office. The distributions differ principally in the intercept, so we may more conveniently characterise the new distribution for two windows by the shift $\Delta\theta_{50}$ in characteristic temperature: considering all individual offices, we obtain $\Delta\theta_{50} = 13.2°\mathrm{C}$. Another possibility is to use ordinal logistic regression (see Section 3.2.6).

We may integrate in the Markov model the possibility to act on multiple windows by adding other possible transitions. For n windows, transition probabilities P_{ij} from i to j windows open ($0 \leq i, j \leq n$) are arranged in an $(n+1) \times (n+1)$ matrix. For instance, with three simulated windows and two currently open, the closing of one or both of open windows is modelled by P_{21} or P_{20} in the case of both and a further opening by P_{23}, with $P_{22} = 1 - P_{20} - P_{21} - P_{23}$. These transitions probabilities may be included in a 4×4 matrix containing 12 independent elements:

$$P_{ij} = \begin{pmatrix} P_{00} = 1 - \sum_{k \neq 0} P_{0k} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} = 1 - \sum_{k \neq 1} P_{1k} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} = \ldots & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} = \ldots \end{pmatrix}$$

This requires the inference of $n \cdot (n+1)$ probability distributions for each occupancy transition. We may similarly derive survival curves for delays until actions on additional windows.

We do not attempt to examine here the probability for two windows to be open in offices with more than one occupant, as individual occupancies are determinant but unknown.

4.3 Discussion

Post hoc ergo propter hoc.
Correlation does not imply causation.

4.3.1 Summary

From the development of a Markov chain model predicting explicit actions on windows, we observed that indoor conditions describe opening actions better than do outdoor conditions – this being our interaction stimulus. But closing actions tend to be better described by outdoor conditions, based on perceived draughts or a risk of overheating when $\theta_{\mathrm{out}} > \theta_{\mathrm{in}}$; likewise whether windows will be left open overnight for cooling purposes. Therefore if we consider the aggregate dataset it is understandable that for a univariate static probabilistic model θ_{out} is statistically stronger than θ_{in} (Section 4.2.2), but this does not make it a better model. This is partly because the previously mentioned subtleties are ignored and partly because, as noted earlier, when using θ_{out} alone the predicted window states are independent of the design of the building; so that occupants of very different adjacent buildings (eg. with minimal and high façade glazing ratio) would be predicted to interact with their windows with similar probability. For such a hypothetical building θ_{out} may again be the best predictor for the (aggregate) logistic model, but with drastically different parameters; in particular $\theta_{\mathrm{out},50}$ is expected to be lower in the highly glazed case.

Any model based on θ_{out} only is thus strongly building-dependent and without generality, requiring separate calibration for each building to which it is applied – an impossible task. Furthermore, it cannot be excluded that the static probabilistic model with both θ_{in} and θ_{out} could be dampened by this lack of generality. On the other hand, although such a model with just θ_{in} is expected to describe occupants' actions with more generality, this also misses important subtleties, particularly in relation to the closing of windows, which undermines its predictive accuracy.

The obtained action probabilities for the discrete-time Markov process (Section 4.2.3) solve this problem, as they directly link the probability for an occupant to take action with the direct environmental stimulus (θ_{in}), whilst also accounting for the fact that θ_{out} has a determinant influence on intermediate closing probability (the sole situation where θ_{out} has a direct connection to the occupant). A possible lack of generality is that the closing probabilities which depend on θ_{out} are likely to depend on window size and opening angle, possibly needing further calibration according to these parameters. The same remarks apply to our continuous-time random process (Section 4.2.4).

In summary then not only do the presented models improve the quality of predictions; they also account for the real stimuli motivating adaptive actions, so improving upon their generality.

4.3.2 Cross-validation

In this work we have presented models of occupants' interactions with windows based on three different methods. We describe here our validation procedure to perform a consistent evaluation of their predictive powers. In addition to this, we will also compare the results from these models with a Bernoulli random variable with constant probability (a random guess based on observed overall opening proportion $p = 0.308$, see Table 2.5), with previous published work (the two versions of the Humphreys algorithm [70, 77], see Section 4.1.1) and with variants based on our models (eg. a discrete-time Markov process with partial sub-models for departure[9]) and a hybrid model (a Markov process with opening durations predicted by a continuous process).

We assess the predictive power of the models by checking four aspects:

- **Discrimination.** Does the model reproduce well the list of observed window states?
- **Overall prediction.** Does the model predict a consistent overall opening ratio throughout the simulation period?
- **Dynamics.** Does the model predict consistent number of actions and delays between actions?
- **Aggregated results.** Is the predicted total number of open windows consistent with observations?

Based on these criteria we will retain the best performing model.

For this validation exercise we have performed 20 repeated simulations using 5 minutes time steps for the whole period with available measurements for the 14 measured offices, producing $20 \times 14 = 280$ sets of simulated window states $W_{\text{sim}}(t)$, to be compared with 14 sets of observed states $W_{\text{obs}}(t)$. This procedure was repeated for each of our models and their variants as well as for the Humphreys algorithm[10].

To avoid any bias in the validation process, for all variants we withdraw one year of measurements (which then forms the validation dataset) and infer regression parameters from the data of the remaining years (the training dataset). Based on these parameters,

[9] This allows a comparison with the approach taken by Yun and Steemers [75, 76].

[10] We used the logistic model based on θ_{in} and θ_{out} adjusted to our data (Section 4.2.2) to simulate both versions of the Humphreys algorithm.

Model	TPR	FPR	ACC	Prop. open	Actions	Dur. open	Dur. closed	Quartiles of error
Exact	100.0%	0.0%	100.0%	30.6%	2.409	115	1275	+0.000 +0.000 +0.000
Bernoulli processes								
- Random guess ($p = r_{\text{open}}$)	30.2%	30.2%	58.7%	30.2%	116.5	5	15	−1.810 **+2.048** +3.714
- Logistic model with θ_{in}	34.0%	29.4%	59.2%	30.7%	111.1	5	15	−0.952 **+1.952** +3.333
- Logistic model with θ_{out}	43.8%	24.8%	65.2%	30.4%	96.40	5	10	−0.143 **+1.286** +2.286
- Logistic model with θ_{in} and θ_{out}	43.8%	24.6%	65.3%	30.4%	96.13	5	10	−0.381 **+1.048** +2.095
- Logistic model with polyn. θ_{out}	45.2%	24.3%	66.0%	30.6%	93.34	5	10	−0.190 **+1.095** +2.048
Markov processes								
- With all variables	38.6%	17.6%	68.5%	28.9%	1.522	1125	4510	−1.238 **+0.095** +1.048
- With one variable	10.5%	5.1%	68.9%	6.7%	1.741	90	4550	−4.381 **−1.476** +0.143
- Close at long departures	24.1%	11.6%	68.3%	15.2%	1.830	190	5235	−4.333 **−1.381** +0.095
- Close at all departures	11.5%	5.2%	69.0%	7.0%	2.086	40	4240	−5.667 **−1.905** +0.000
Continuous processes								
- Weibull distributions	28.8%	19.0%	64.8%	22.1%	112.6	5	5	−0.810 **+1.476** +2.905
- Hybrid model	43.2%	16.1%	71.3%	24.3%	1.463	185	4815	−0.952 **+0.048** +0.905
Humphreys algorithm								
- Version 2007	30.1%	32.8%	56.6%	31.8%	0.181	14040	20680	−2.238 **+1.762** +4.238
- Version 2008	18.5%	13.9%	65.7%	15.1%	0.012	153700	622400	−3.429 **−0.286** +0.857
Deterministic models								
- Always closed	0.0%	0.0%	69.8%	0.0%	0	NA	∞	−6.000 **−2.000** +0.000
- Always open	100.0%	100.0%	30.2%	100.0%	0	∞	NA	+8.000 **+12.000** +14.000

Table 4.9: Validation parameters: true positive rate, false positive rate, accuracy, total proportion of simulated time steps with window open, average number of opening actions per day, median duration (min.) of openings and closings and quartiles of the distribution of the error on total number of windows open

we simulate the window states at the validation set and repeat this procedure until all the data have been selected once as validation sets. With eight years of measurements, this corresponds to an eight-fold cross-validation. This procedure was preferred to random sub-sampling validation as it ensures that representative subsets of seasonal conditions are used as training sets.

Discrimination

A general validation procedure should involve comparing each model's ability to directly reproduce observed window states. We would thus obtain results that may be classified in four groups, as mentioned in Section 3.2.3: a predicted open window is (i) truly open (TP), (ii) falsely open (FP, Type I error); a predicted closed window is (iii) truly closed (TN), (iv) falsely closed (FN, Type II error). We present these proportions for each simulated model in Figure 4.12. Based on the definitions provided in Section 3.2.3, we may then accumulate these results to define the overall true positive rate (TPR), the false positive rate (FPR) and the *accuracy* $\text{ACC} = (\text{TP} + \text{TN})/(\text{P} + \text{N})$, which gives the proportion of correct predictions.

Based on the twenty repeated simulations, these indicators are computed and displayed in Table 4.9. Applying the concepts introduced in Section 3.2.3, we may also draw the corresponding points in the receiver-operating space (Figures 4.13(a)-4.13(b)), in which our ideal model would be located at minimum x and maximum y. Each small point corresponds to a single simulation of an office, while parameters referring to the aggregated simulation results of a model are plotted as bigger solid points.

As expected from our previous statistical tests, a Bernoulli random variable based on the univariate logistic model with θ_{in} performs much worse than with θ_{out}. The distribution with two variables discriminates slightly better compared with θ_{out} alone, while the

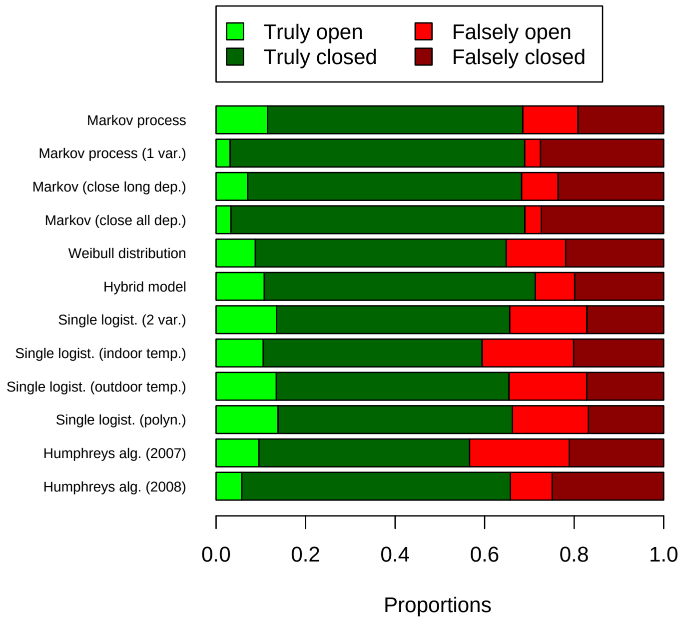


Figure 4.12: Classification of simulation results

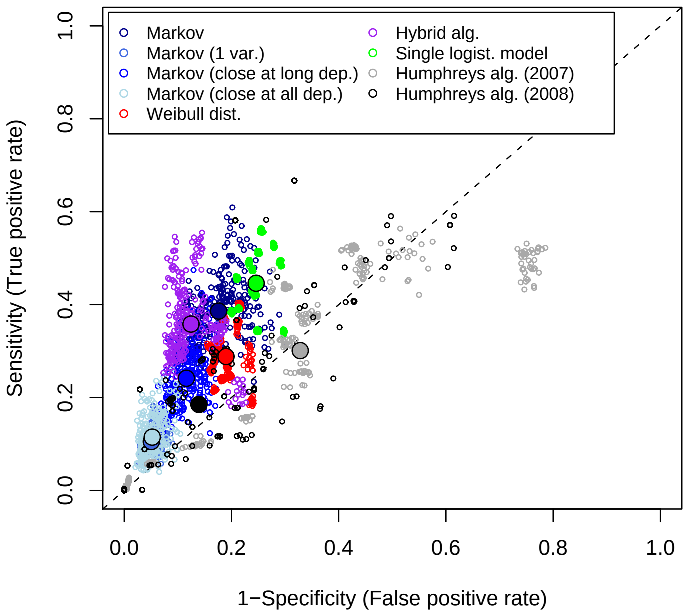
polynomial logit with θ_{out} offers best discrimination among the tested Bernoulli processes (Figure 4.13(b)).

The discrete-time Markov process gives lower values of TPR and FPR. This suggests that this model is more "conservative", that is it predicts less openings than the Bernoulli processes (and misses slightly more of them), but on the other hand much less false openings are predicted. Furthermore, it has higher overall accuracy. We can also observe that the quality of predictions decreases drastically if we attempt to treat actions on departure simplistically.

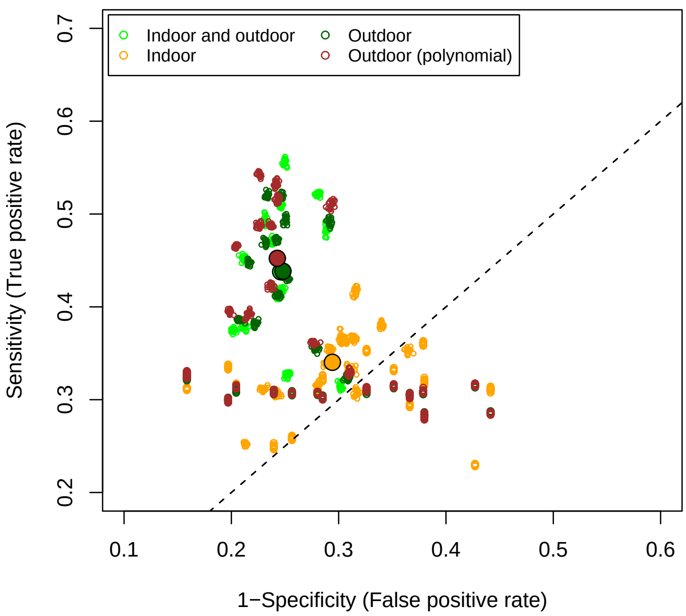
The Weibull distribution for the continuous-time random process is generally slightly less "conservative" than the Markov process, albeit with a slightly lower accuracy. However, this model is computationally much faster. In an attempt to find a good compromise between accuracy and speed we have therefore developed a hybrid model based on the discrete-time Markov process, using a Weibull distribution for opening durations only. Our simulations show that this provides the highest accuracy, while increasing TPR and reducing FPR compared with the plain Markov model. According to the discrimination criteria, this model offers the best performance.

Overall window opening ratio

Based on the total presence duration $T_{\text{pres,tot}}$ and the total window opening time $T_{\text{open,tot}}$, we define for each office the overall window opening ratio as $r_{\text{open}} = T_{\text{open,tot}}/T_{\text{pres,tot}}$. We show observed and predicted values in Table 4.9 and Figure 4.14. Overall opening ratios predicted by the Markov processes are rather low, particularly when night ventilation behaviour is neglected. The Bernoulli processes generally reproduce well this parameter,



(a) Discrete-time Markov processes, continuous-time model and their hybrid model



(b) Bernoulli processes based on logistic models

Figure 4.13: Cross-validation results: simulation results in the receiver-operating characteristic space

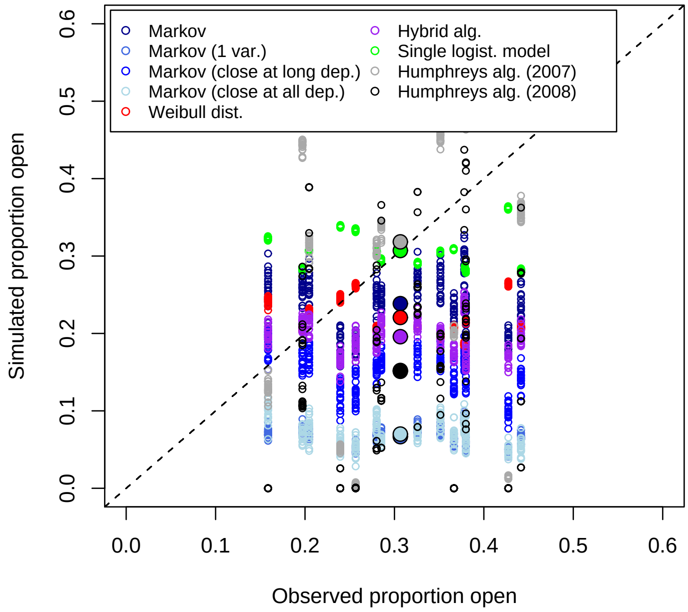


Figure 4.14: Observed and simulated proportion of windows open

likewise the Humphreys algorithm. However the Markov process predicts the greatest variability between results.

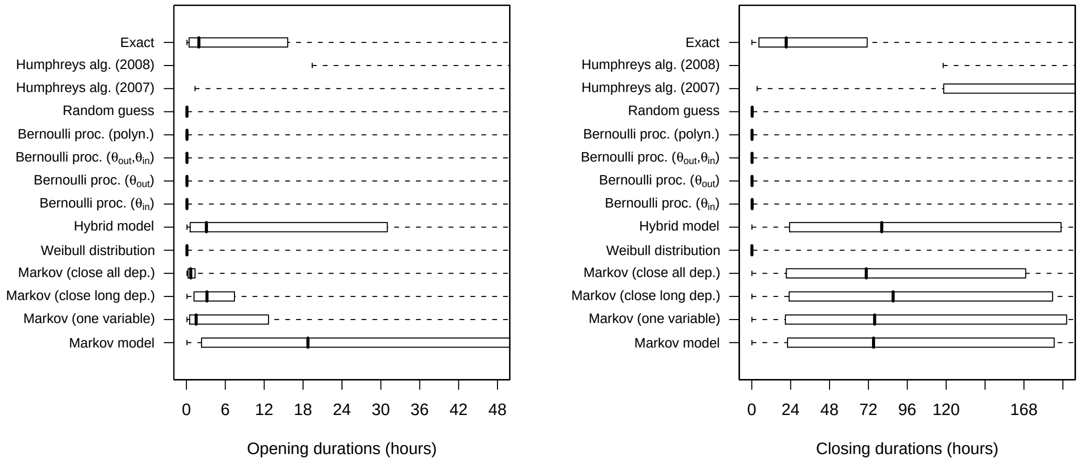
It is worth noting that all the models predict similar total opening ratios between simulated offices, which shows that the added refinements do not reproduce this variability between occupants.

Number of actions, opening and closing median durations

As expected, the models based on Bernoulli random variables do not predict coherent delays between actions (Table 4.9), as they are not explicitly based on any dynamics in their formulation. The Markov model overestimates these durations; that is it does not predict enough actions from occupants. The hybrid model best reproduces observed opening durations, while predicting a coherent number of actions. As with the Markov model, the Humphreys algorithm predicts durations that are too long, which may be caused by too large a deadband (indeed a significant proportion of actions occurs at moderate temperatures) or by ignoring the determinant influence of occupancy transitions.

Aggregated results

We extend here our validation approach to the global simulation results and check whether the models are able to reproduce observed window openings on a large scale. Based on individual simulations of window states in the 14 offices, we compute the total number of windows open $N_{\mathrm{sim}}(t) = \sum_{k=1}^{14} W_k(t)$ for each simulation. Averaging over the 20 sets of $N_{\mathrm{sim}}(t)$ we deduce the mean predicted number of windows open at each time step $\bar{N}_{\mathrm{sim}}(t)$.



(a) Opening durations (b) Closing durations

Figure 4.15: Box-and-whisker plots of observed and simulated opening and closing durations

Comparing with the observed number of windows open $N_{\text{obs}}(t)$, we may compute the error $\varepsilon(t) = \bar{N}_{\text{sim}}(t) - N_{\text{obs}}(t)$ at each time step.

We show in Table 4.9 the quartiles of this error for each model. From this we observe that the hybrid model offers the best performance on aggregated results, with smallest error magnitude. Furthermore, comparisons of measured and predicted numbers of windows open (Figures 4.16-4.17) indicate that this model reproduces well the temporal variation of window openings[11].

Overall recommendation

From these validation results, we recommend the use of a hybridised model including a discrete-time Markov model for the prediction of openings and a Weibull distribution for their duration. Although, this model somewhat underestimates the overall opening ratio, we observed that it offers the highest accuracy, produces the best discrimination between window states, reproduces acceptably the delays between actions and offers the most reliable aggregated predictions at the scale of a whole building. It also predicts realistic opening durations that are not constrained by the choice of the time step – an additional advantage that the cross-validation process did not underline. A detailed description of this algorithm for implementation in building simulation tools is provided in Section 7.1.

We may expect improved predictive accuracy when this algorithm is applied in building simulation tools, where it is coupled with an air flow model: indeed our validation procedure is based on observed temperatures, which do not evolve under the influence of simulated actions, likewise probabilities of further actions.

[11] This refers to the period from 27 January 2005 to 14 January 2006, which offers representative climatic conditions and uninterrupted measurements.

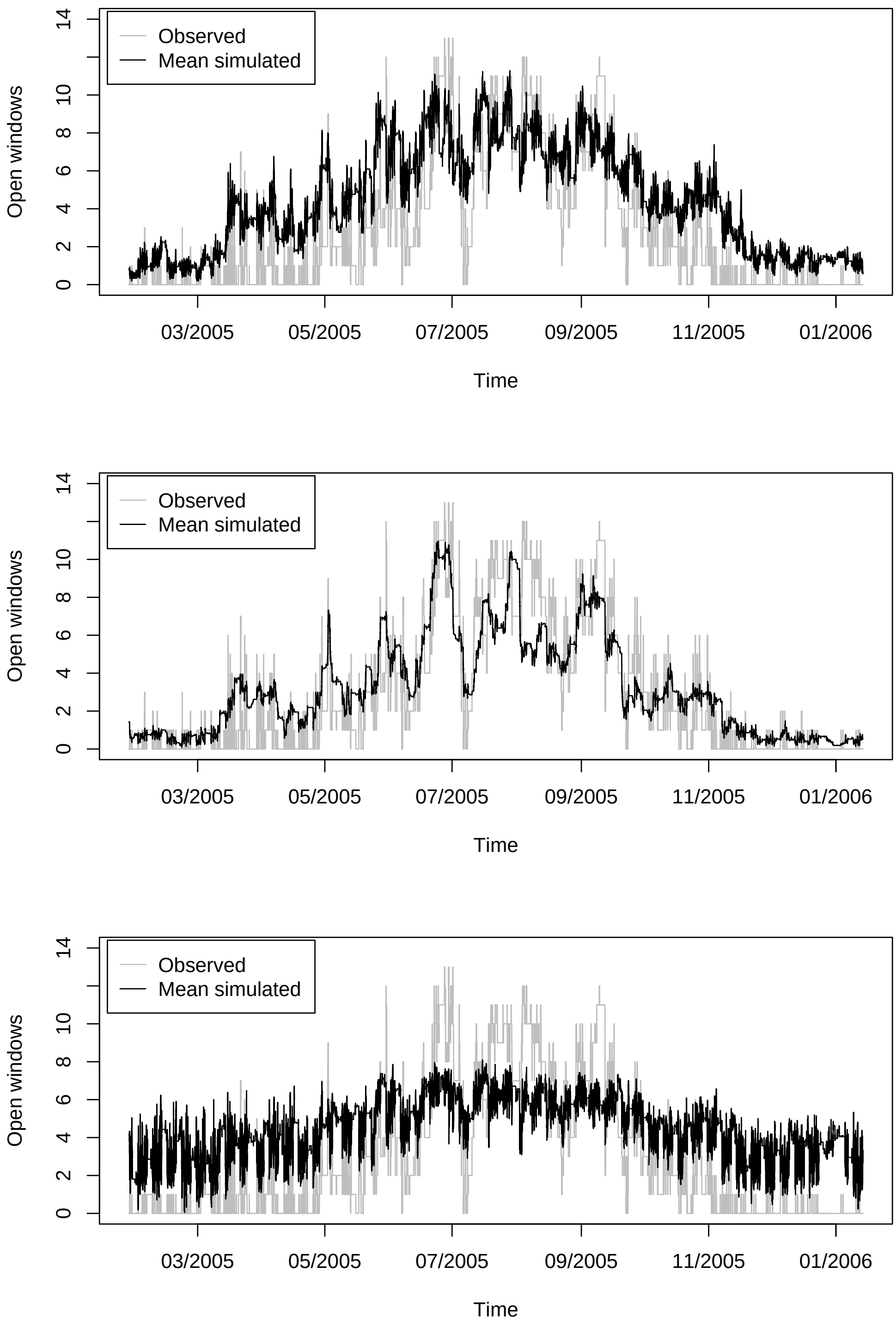


Figure 4.16: Observed and mean simulated number of windows open for 5 minute time steps on a period of a year using the Bernoulli process based on θ_{out} and θ_{in} (top), the Markov model (middle) and the continuous-time process (bottom)

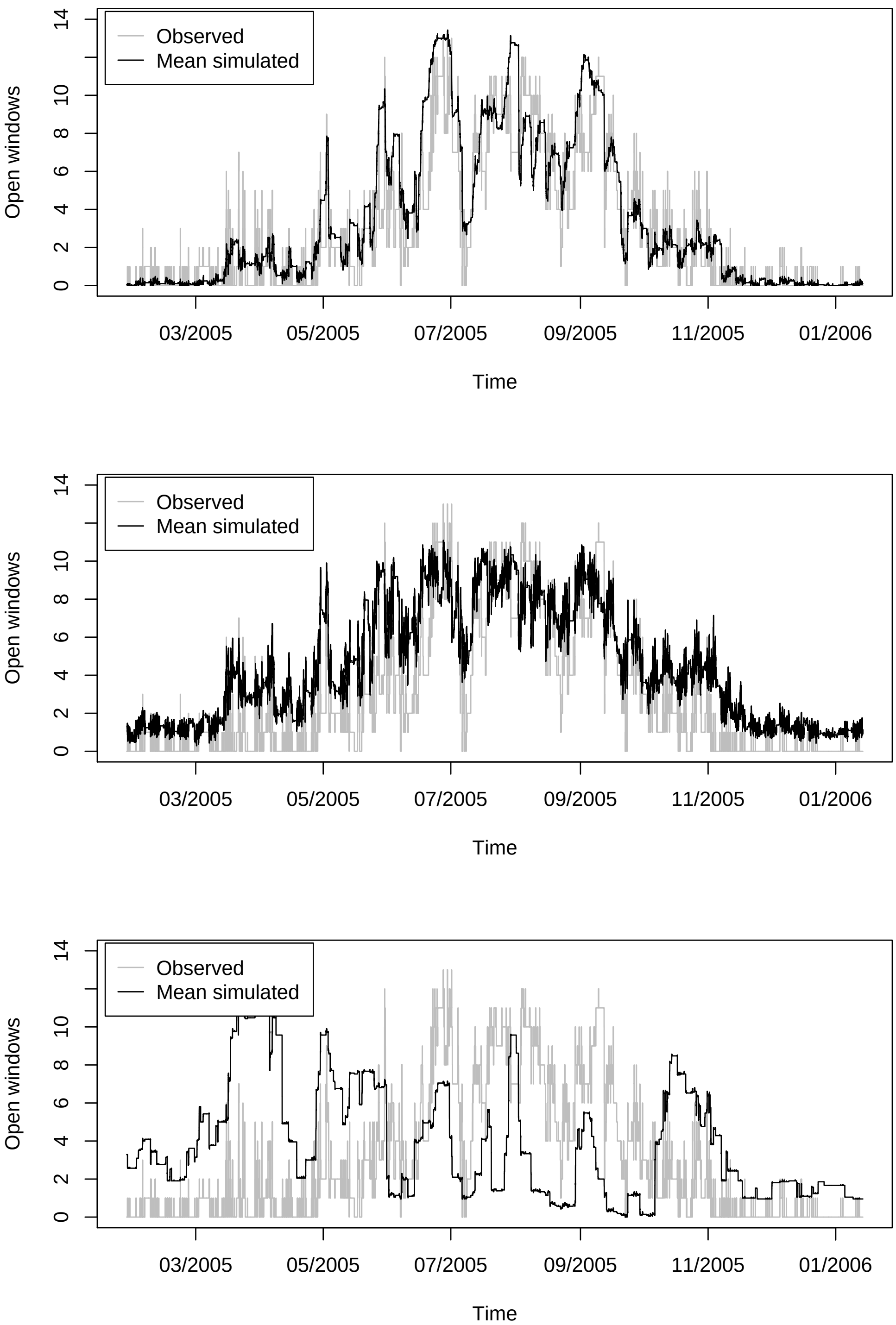


Figure 4.17: Observed and mean simulated number of windows open for 5 minute time steps on a period of a year using the hybrid model (top), the polynomial logistic model (middle) and the Humphreys algorithm, version 2007 (bottom)

4.3.3 Treating opening angles

We do not treat in this work the behaviour of occupants regarding window opening angles as this data was not recorded during our field surveys. We however do propose two possible ways to account for them.

A simple approach might involve a sub-model to predict tilting or axial opening from closed to open, based on relevant environmental parameters. If an axial opening is performed, the opening angle ϕ may be drawn (using the inverse function method) from a probability distribution, which we expect to be mainly dependent of θ_{out}, window size and possibly v_{wind} and α_{wind}. If the window may be tilted, a preliminary algorithm chooses between tilting and axial opening.

This approach neglects the possibility to vary the angle whilst the window is open. If this needs to be accounted for, an approach similar to that of Fritsch et al. [59] would seem to be a sensible solution. This implies the definition of transition probabilities $P_{\alpha,\beta}$ from angle α to β, depending on the most influential parameters.

4.4 Conclusion

Based on almost eight years of observations we have developed three different modelling methods for the prediction of actions on windows: an inhomogeneous Bernoulli process based on a logistic probability model (Section 4.2.2), a discrete-time Markov process with sub-models for different occupancy statuses (Section 4.2.3) and extended this latter to a continuous-time random process (Section 4.2.4). Supported by rigorous cross-validation, we have demonstrated the superiority of a discrete-time Markov process approach and its strong added value compared with existing models (Section 4.3). We have furthermore inferred a continuous-time model that could be efficiently used for a fast calculation of opening and closing durations.

We have finally tested possible combinations in these approaches and selected a hybrid model. This hybrid combines the accuracy of the discrete-time Markov process with the efficiency of the continuous-time model for opening durations. For this we also describe in Section 7.1 a step-by-step process by which the algorithm may be implemented.

We have also studied the diversity in individual behaviours and described a possible method to integrate them if necessary; likewise a method to integrate in the Markov model the possibility of acting on multiple windows.

However, there remain some outstanding issues to be addressed in the modelling of window opening and closing behaviour; particularly in respect to the angle of window opening as opposed to simply the state open and closed. Finally we are mindful that, although we have demonstrated the validity of the formulation of our proposed new algorithm, the parameters for its calibration are strictly speaking currently limited to just one building.

It would thus be desirable to make use of measurements from other buildings (residential in particular), in which opening angles are also recorded, to have a stronger basis for calibration and thus application to other simulated buildings. Such surveys might also usefully include other variables which may influence actions on windows, such as radiant temperature or indoor relative humidity (particularly for tropical climates). Factors related to indoor air quality (eg. CO_2 or pollutant concentration) should also be treated; however it is plausible that the inclusion of T_{pres} in our intermediate openings model ($P_{01,\text{int}}$) could implicitly account for this (at least in part).

Finally, although they are based on cross-validation, the values of the indicators of

Table 4.9 strictly refer to the performance of the algorithms based on data from the building where observations are validated. They allow for a comparison of the predictive powers of the variants tested, but they do not necessarily estimate the accuracy of the algorithms if applied to other buildings. Simulation and comparison with data from other surveys (eg. using data from one building as training set and perform validation on data from another building) are necessary to reliably estimate the accuracy of these algorithms for their application to other situations.

Chapter 5

Modelling actions on shading devices

Based on seven years of continuous measurements (Section 2.1), we have analysed in detail the occupancy, thermal and visual parameters influencing actions on shading devices in order to derive an accurate model for the prediction of their usage in office buildings. This chapter begins by presenting some of the key findings from these analyses. Informed by other developments in the literature, we go on to propose an approach for a comprehensive stochastic model for simulating blind usage. This model is based on a Markov process taking rigorously selected predictors (initial blind status, indoor and outdoor illuminance) as input variables to predict lowering and raising actions performed by occupants. A separate sub-model then predicts the chosen shaded fraction. An assessment of the predictive accuracy of simulations is then presented for several modelling variants using our measured data, from which the best performing model variant is selected. Based on local visual stimuli, the form of this model is expected to be directly applicable to other buildings. Finally, individuals' behaviours are examined and a possible approach for modelling behavioural diversity is discussed.[1]

5.1 Introduction

Shading devices play a central role in the heat gains of a building and therefore on its energy performance. It is thus useful to predict their use by occupants, particularly where automatic controls are lacking, in dynamic building thermal simulation tools, in order to correctly assess the availability of daylight, overheating risk and the visual comfort of occupants. We present in this introduction a short summary of previous research in this domain and outline the need for further developments.

5.1.1 State of the art

Past research in the domain of occupants' actions on blinds was based on two motivations: first, the development of control algorithms to allow automated systems to adjust shading in order to optimise solar heat gains and visual comfort; second, the prediction of actions performed by occupants in order to integrate them into building simulation tools. We are interested here in the latter approach. To this end we briefly review here previously published findings.

[1] A substantial proportion of this chapter was presented at the 11$^{\text{th}}$ International Building Performance Simulation Association Conference [89]. An extended version has also been accepted for publication in the *Journal of Building Performance Simulation* [90].

Based on analysis of variance of two months' measurements in a single building, Rea [91] observed that blind occlusion varied significantly between different sky conditions (cloudy or clear), the building orientation (east, south and west) and the interactions between the levels of these latter variables. He noticed that occupants made little attempt to change blind position during the day.

From measurements on four buildings, Inoue at al. [92] noticed that the frequency of blind usage varied with orientation and weather conditions and that it was very particular to the building surveyed. They concluded that if the direct solar radiation on a façade exceeded some value between 12 and 58 W/m^2, blind occlusion is then proportional to sunlight penetration depth.

Reinhart [78, 79] developed the Lightswitch-2002 algorithm, based on a review of studies in several countries, which dynamically models manual and automatic control of blinds and lights on a 5 minutes time step, and was integrated into ESP-r [20]. This model distinguishes two types of behaviour towards blinds use: dynamic (adjusted on a daily basis) and static (permanently lowered). For this blinds are lowered if the irradiance on the workplace reaches the threshold of 50 W/m^2; they are otherwise kept open. Lightswitch-2002 appears to be the first attempt to develop a formal algorithm for the prediction of actions on blinds. It does nevertheless have some limitations: it predicts that blinds are opened only once a day and it uses a rigid threshold for visual comfort.

Nicol and Humphreys [68] and Haldi and Robinson [74] mentioned an increase in the proportion of blinds lowered as indoor (and outdoor) temperature rises. These former go on to suggest that the effect seems marginal and that it may simply reformulate the effect of a primary variable linked to visual stimuli. Instead, they recommend the use of outdoor illuminance as the explanatory variable.

In their pilot study of eight offices Sutter et al. [93] observed that occupants mostly set their blinds fully raised or lowered. They also reported an "hysteresis phenomenon" in the use of blinds; that is the illuminance level at which occupants lower their blinds is higher than that at which they raise them. It was noticed that most occupants keep their blinds down until the illuminance is very low, before raising them. They observed that a logistic function (see Section 3.2) – with the logarithm of external vertical global illuminance as driving variable - fits well the percentage of blinds raised. The possibility of an independent effect of temperature was also suggested.

Using Bayesian analysis, Lindelöf and Morel [94] analysed actions on lighting and blinds using data from the LESO building (see Section 2.1.1) to infer a probability distribution of visual discomfort that reaches a minimum for horizontal workplane illuminance of 800 to 1200 lux.

Mahdavi et al. [95, 96, 97] observed from a field survey in three buildings, that actions on shading devices occurred on average once every week, with significant differences between occupants.

Finally, based on measurements in two air-conditioned buildings, Inkarojrit [98] tested a model formulated as logistic probability distributions, with a range of different parameters; retaining four predictors: average luminance of the window, maximum luminance of the window, vertical solar radiation and self-reported sensitivity to brightness. The experimental design did not however support the development of a comprehensive model, as only the behaviour on arrival is studied.

Actions	Arrival	Intermediate	Total	Full	Partial
Lower blinds					
Lowering	2105 (2.3%)	3624 (0.33%)	5729 (0.48%)	10.5%	89.5%
Raising	1297 (1.4%)	3541 (0.32%)	4838 (0.40%)	66.8%	33.2%
No action	88928 (96.3%)	1103967 (99.45%)	1192895 (99.12%)		
Upper blinds					
Lowering	2308 (2.5%)	4139 (0.37%)	6447 (0.53%)	45.6%	54.4%
Raising	1495 (1.6%)	4061 (0.37%)	5556 (0.46%)	70.3%	29.7%
No action	88527 (95.9%)	1102932 (99.26%)	1191459 (99.01%)		
Total	92330 (7.7%)	1111132 (92.3%)	1203462 (100%)		

Table 5.1: Classification of observed actions on lower and upper blinds with respect to occupancy status (2nd to 4th columns) and the proportion of actions to adjust blinds to their fully (un)shaded fractions (5th and 6th columns)

5.1.2 Perspectives

This short review underlines the need for further research in order to correctly integrate occupants' behaviour with respect to blinds, as the majority of published studies have not been supported by the data required to infer comprehensive models (or the opportunity for doing so has not been taken) accounting for the range of possible explanatory variables and how behaviour might vary at arrival, during occupancy and upon departure. In order to correctly predict occupant behaviour we use a rigorous statistical methodology to select the relevant driving variables for actions. In this chapter we also propose an algorithm for application in dynamic building simulation tools, with a higher degree of realism than is the case with the deterministic inputs of Lightswitch-2002.

5.2 Patterns of actions on blinds

We present briefly in this section several preliminary observations which should provide useful guidance for the choice of an appropriate modelling approach. The statistical software R [86] was used for all data analyses and for the programming of models.

Two offices (201 and 202) have a very particular configuration of blinds and so were removed from the database. A statistical summary of all the relevant variables measured during occupied periods is presented in Table 2.4 and Figure 2.6(b).

5.2.1 Actions and occupancy-related effects

Table 5.1 shows an overview of actions performed by occupants with respect to occupancy status, based on five minute time steps. It can be noticed that occupants adjust their blinds more often on arrival than during their presence, up to 5.7 times more often for lower blinds and 5.5 times more often for upper blinds. A sensible model should thus account for these differentiated action probabilities. On the contrary, there is no significant increase in action rate when occupants leave their offices. Actions at departure are thus merged with intermediate actions in Table 5.1. Occupants do not seem therefore to adjust their blinds for predictive purposes, for instance to prevent excess solar gains during their absence. This fact may advocate for the use of an automatic controller to optimise heat gains during occupants' absence. Finally, we do not observe any significant differences in occupancy-related behaviour between offices and between floors.

Actions outside the arrival period are extremely rare, occurring during less than a one hundredth of our 5 minute time steps. However, it is not yet clear whether these actions at arrival are an intrinsic effect of occupancy transition (eg. the occupant perceives a sudden difference in the visual environment upon arrival, motivating action), or whether it is more likely that the blind position is inadequate after an unoccupied period due to climatic changes.

Table 5.1 also reveals that lowering actions are more frequent than raising. This pattern is linked to the fact that lowering blinds to their fully shaded position are much less commonly performed than full raising actions (Table 5.1, 5$^{\text{th}}$ and 6$^{\text{th}}$ columns). This shows that occupants set their blinds more carefully when lowering them (which results in repeated partial lowering actions), where care is taken to maintain view and a suitable internal illuminance, than when raising them (where are less perturbed by surplus illumination).

It can also be noticed that upper blinds are slightly more frequently used. A likely explanation lies in the fact that lowered upper blinds leave occupants' view almost unobstructed, and so they are preferred as a first choice to reduce glare.

5.2.2 Observed shaded fractions

A histogram of the prevailing unshaded fractions is given in Figures 5.1(a)-5.1(b). We observe that lower blinds are in a fully raised position 67.1% of occupied time and fully lowered 5.2% (56.4% and 21.7% for upper blinds), extreme unshaded fractions are thus overrepresented (around 75% of occupied periods). This pattern may be due to the fact that blinds are set in movement by pressing a command, while another press is needed for stopping it. We may expect a different behaviour for other types of command, such as crank-operated shading devices.

Upper blinds are four times more likely to be fully lowered; partly because these blinds do not obstruct the view, but also because of the efficiency of the anidolic reflector in redirecting external illumination.

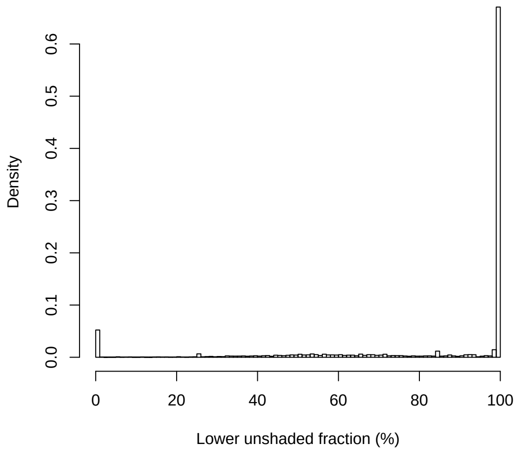
The chosen fractions display small variations among the surveyed offices. The occurrence of fully raised lower blinds exceeds 75% in offices 104, 105 and 106 (Figure 5.1(c)), which may be a consequence of the greater distance between occupants' desktop and the external façade (so that greater proportion of the sky vault is directly visible).

5.2.3 Variables influencing the state of blinds

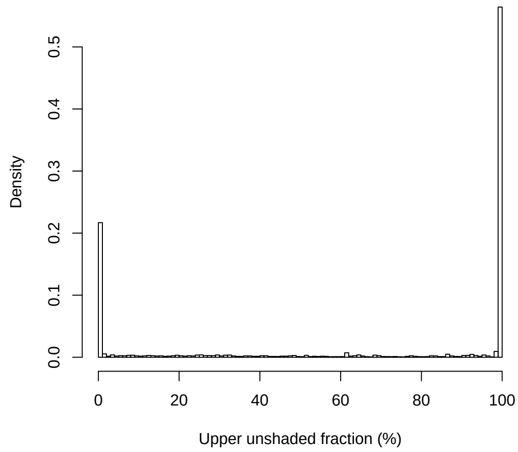
In order to develop a predictive model for blind position, it is of interest to study the variation of the above shaded fractions in conjunction with key environmental variables. To this end we show in Figure 5.2 the repartition of lower and upper blind positions as functions of θ_{in}, θ_{out}, E_{in} and $E_{\text{gl,hor}}$.

The state of lower blinds seems not to be directly linked with any of these variables. However clear increases in upper shaded fraction are noticeable when θ_{in}, θ_{out} and $E_{\text{gl,hor}}$ rise.

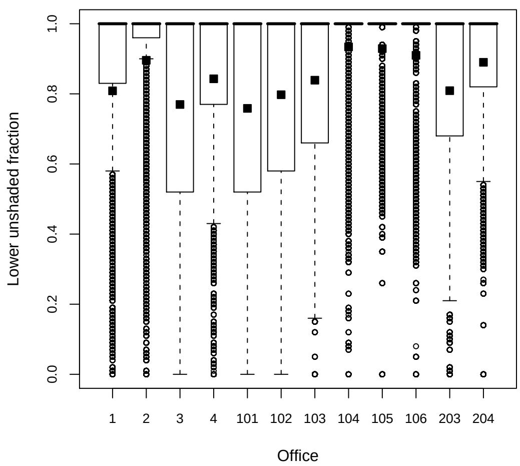
An interesting approach to infer a distribution predicting the state of blinds with respect to their unshaded fraction is to perform ordinal logistic regression (see Section 3.2.6), which is appropriate for non-binary outcomes, as a shaded fraction can take any value between 0



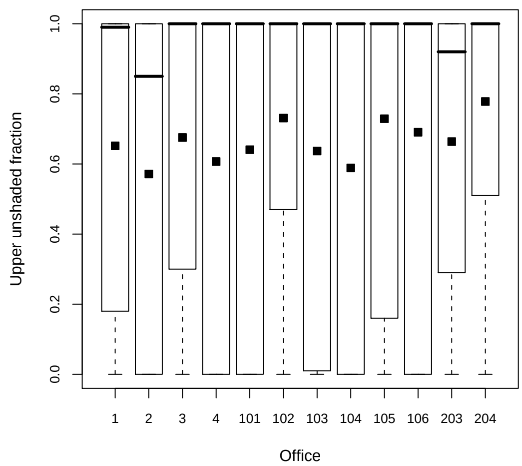
(a) Lower blinds: histogram of observed unshaded fractions



(b) Upper blinds: histogram of observed unshaded fractions



(c) Lower blinds: box-and-whisker plots of observed unshaded fractions per office



(d) Upper blinds: box-and-whisker plots of observed unshaded fractions per office

Figure 5.1: Prevailing shaded fractions in the surveyed offices for the whole occupied period. Box-and-whisker plots show the 1st and 3rd quartiles of unshaded fractions as boxes, the median fraction as a thick line and the most extreme point inside 1.5 times the interquartile range as whiskers beyond which we have outliers, and solid squares for means.

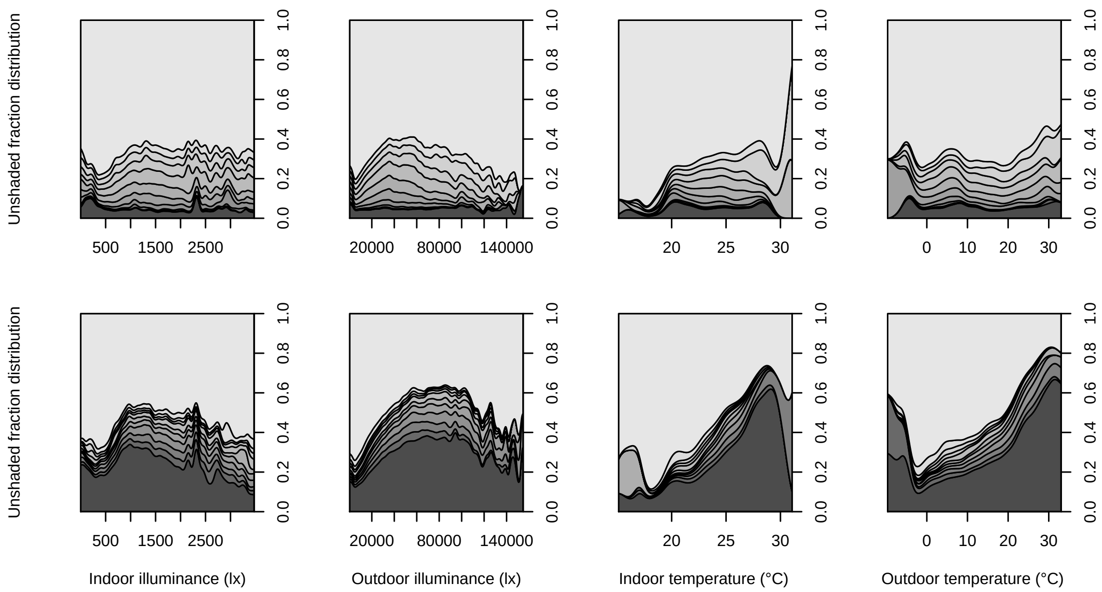


Figure 5.2: Plots of probability distributions for lower (top) and upper (bottom) unshaded fractions, conditional on E_{in}, $E_{\mathrm{gl,hor}}$, θ_{in} and θ_{out}. Observed prevalence of unshaded fractions is shown as a gray scale ranging from fully lowered (dark gray) to fully raised (light gray).

and 1. For this the proportional odds model gives a probability for the unshaded fraction B to be at least a fraction B_j as the function:

$$p(B \geq B_j | x_1, \ldots, x_n) = \frac{\exp(a_j + \sum_{i=1}^{n} b_i x_i)}{1 + \exp(a_j + \sum_{i=1}^{n} b_i x_i)} \tag{5.1}$$

where B_j may be set to any unshaded fraction between 0 and 1. With this convention, we have a regression parameter b_i per predictor x_i and an intercept a_j per threshold shaded fraction B_j. We have carried out this procedure with several variables of interest, without convincing results, as the quality of adjustment is low (see Table 5.3). Although the obtained distributions summarise, in a single formula, the position of blinds with respect to chosen predictors, they do not describe the dynamics of actions and their application with Monte-Carlo simulation is not straightforward, implying sampling from a multinomial distribution. It is thus more appropriate to directly model actions rather than probability distributions for the state of blinds.

5.2.4 Stimuli for action

Before formally inferring a model for the probability of raising and lowering blinds, we studied the state of several variables of interest, when actions occur. We show in Figure 5.3

the kernel density estimators[2] of indoor illuminance, indoor temperature, sun elevation and azimuth at the moment of lowering and raising, to be compared with their baseline distributions on the total presence duration.

From these observations, we notice that the distributions of E_{in} and $E_{\text{gl,hor}}$ strongly differ from the prevaviling conditions during time steps when actions are performed. As expected, low values of these latter variables are associated with raising actions, and conversely for lowering actions. This provides strong evidence for the prevalence of visual stimuli for performing actions on blinds. The distributions of thermal variables θ_{in} and θ_{out} are not significantly modified when actions are taken.

Actions are more often performed at specific sun elevation and azimuth. However, further analysis should assess the existence of an independent effect: for instance, Figure 5.3 (bottom right) shows increased lowering actions in the morning and raising actions in the late afternoon, but this can be an indirect consequence of rising and lowering values of E_{in} and $E_{\text{gl,hor}}$.

Finally, these distributions do not differ strongly between lower and upper blinds. A similar set of driving variables is thus responsible for both these actions. Distributions of other available variables were analysed, producing no noticeable shifts in distributions.

5.3 Model for actions on blinds

In order to account for the real dynamic processes leading occupants to perform actions on blinds, we infer actual probabilities of lowering or raising blinds, provided relevant physical parameters, determined through statistical analysis of observations. Our approach is first to determine the driving variables influencing actions on lower blinds and then to formulate lowering and raising probabilities (Sections 5.3.1-5.3.2). Based on the observed over-representation of actions on arrival, we will distinguish occupancy situations (arrived, intermediate and departing) and check for the significance of their differences. As noted earlier, besides actions we should also model the chosen position of our blinds and not simply whether an action has taken place or not (Section 5.3.3). This reasoning is also applied to the case of upper blinds (Section 5.3.4).

We deduce the action probabilities as logistic models (Section 3.2), where we adopt the following notation:

$$\begin{aligned}\text{logit}(p) = \log(p/(1-p)) = a + b_{\theta_{\text{in}}}\theta_{\text{in}} + b_{\theta_{\text{out}}}\theta_{\text{out}} + b_{E_{\text{in}}}E_{\text{in}} + b_{E_{\text{out}}}E_{\text{gl,hor}} \\ + b_{Ig}I_{\text{gl,hor}} + b_{Id}I_{\text{diff,hor}} + b_{Ib}I_{\text{beam}} + b_{BL}B_L + b_{BU}B_U + b_L f_L,\end{aligned} \tag{5.2}$$

where a and b_i are the regression parameters; as summarised in Table 5.2. Following a similar method as in Chapter 4, starting with univariate models we proceed to consider models with several variables; continuing this procedure to other predictors until no further addition may provide extra significance.

[2]The kernel density estimator based on a sample $X_1, \ldots, X_n$ from distribution f is defined as

$$\hat{f}(x) = \frac{1}{nh}\sum_{j=1}^{n} g\left(\frac{x - X_j}{h}\right),$$

where $h > 0$ is the bandwidth and $g(x)$ is a kernel function (a symmetric probability density with mean zero and unit variance), we used here the standard normal density.

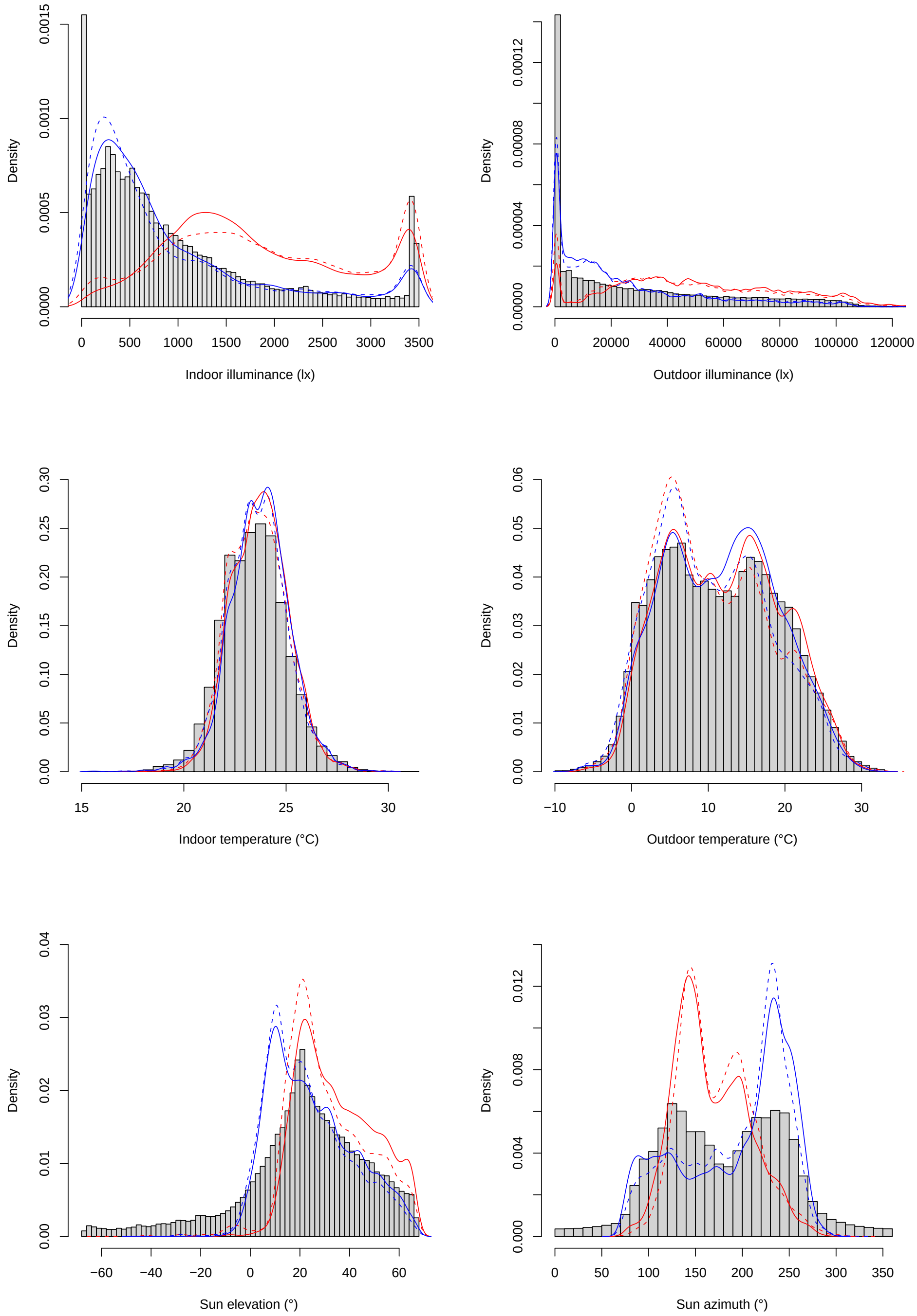


Figure 5.3: Histograms of observed values of E_{in}, $E_{\text{gl,hor}}$, θ_{in}, θ_{out}, ζ and α during the whole measurement period (gray rectangles). Kernel density estimates of the distributions of these variables when lowering (red) and raising (blue) actions were performed are superposed, for lower (solid lines) and upper (dashed lines) blinds.

Type	Parameters	Lower blinds		Upper blinds	
		Estimate	χ^2	Estimate	χ^2
$P_{\text{lower,arr}}$	a	-7.41 ± 0.16		-7.29 ± 0.11	
	$b_{E_{\text{in}}}$	(10.35 ± 0.19)·10^{-4}	3005.04	(9.48 ± 0.21)·10^{-4}	2041.65
	b_{BL}	2.17 ± 0.16	177.79		
	b_{BU}			2.18 ± 0.10	431.97
	$b_{E_{\text{gl,hor}}}$			(6.66 ± 0.76)·10^{-6}	76.73
$P_{\text{raise,arr}}$	a	-1.520 ± 0.051		-1.699 ± 0.041	
	$b_{E_{\text{in}}}$	(-6.54 ± 0.46)·10^{-4}	202.87	(-5.24 ± 0.54)·10^{-4}	92.84
	b_{BL}	-3.139 ± 0.068	2127.15		
	b_{BU}			-3.916 ± 0.094	1738.22
	$b_{E_{\text{gl,hor}}}$			(-21.8 ± 1.3)·10^{-6}	283.34
$P_{\text{lower,int}}$	a	-8.013 ± 0.086		-8.211 ± 0.059	
	$b_{E_{\text{in}}}$	(8.41 ± 0.13)·10^{-4}	4173.20	(8.34 ± 0.14)·10^{-4}	3506.19
	b_{BL}	1.270 ± 0.086	216.84		
	b_{BU}			1.533 ± 0.056	741.96
	$b_{E_{\text{gl,hor}}}$			(5.69 ± 0.53)·10^{-6}	115.4
$P_{\text{raise,int}}$	a	-3.625 ± 0.030		-3.629 ± 0.025	
	$b_{E_{\text{in}}}$	(-2.76 ± 0.22)·10^{-4}	155.24	(-2.90 ± 0.26)·10^{-4}	125.00
	b_{BL}	-2.683 ± 0.040	4600.57		
	b_{BU}			-3.365 ± 0.051	4398.09
	$b_{E_{\text{gl,hor}}}$			(-16.86 ± 0.68)·10^{-6}	622.32
$P_{\text{full lower}}$	a	-0.27 ± 0.14		-0.435 ± 0.097	
	b_{BL}	-2.23 ± 0.16			
	b_{BU}			0.150 ± 0.096	
	$b_{E_{\text{gl,hor}}}$	(0.91 ± 1.33)·10^{-6}		(2.50 ± 0.79)·10^{-6}	
$P_{\text{full raise}}$	a	0.435 ± 0.062		1.543 ± 0.044	
	b_{BL}	1.95 ± 0.11			
	b_{BU}			-0.56 ± 0.10	
	$b_{E_{\text{gl,hor}}}$	(-2.31 ± 0.11)·10^{-5}		(-2.12 ± 0.11)·10^{-5}	

Table 5.2: Regression parameters and analysis of deviance for action probabilities and for full lowering and raising probabilities (in all cases, $p < 0.001$ according to the Wald and likelihood ratio tests)

Model	Lower blinds				Upper blinds			
	AUC	$\mathbf{R_N^2}$	**B**	$\mathbf{D_{xy}}$	**AUC**	$\mathbf{R_N^2}$	**B**	$\mathbf{D_{xy}}$
$P(B > B_j \| \theta_{\text{in}})$	0.519	0.002	0.220	0.039	0.604	0.050	0.236	0.207
$P(B > B_j \| \theta_{\text{out}})$	0.509	<0.001	0.221	0.018	0.616	0.064	0.233	0.232
$P(B > B_j \| E_{\text{in}})$	0.530	0.003	0.220	0.060	0.541	0.001	0.245	0.081
$P(B > B_j \| E_{\text{gl,hor}})$	0.541	0.005	0.219	0.081	0.629	0.074	0.229	0.258
$P(B > B_j \| E_{\text{in}}, E_{\text{gl,hor}})$	0.629	0.074	0.229	0.258	0.648	0.092	0.226	0.295
$P_{\text{lower,arr}}$	0.833	0.191	0.021	0.665	0.865	0.232	0.023	0.730
$P_{\text{raise,arr}}$	0.881	0.172	0.014	0.762	0.893	0.256	0.015	0.786
$P_{\text{lower,int}}$	0.778	0.081	0.003	0.555	0.821	0.118	0.004	0.642
$P_{\text{raise,int}}$	0.861	0.091	0.003	0.722	0.841	0.142	0.004	0.682
$P_{\text{full lower}}$	0.564	0.051	0.090	0.128	0.527	0.002	0.248	0.054
$P_{\text{full raise}}$	0.715	0.162	0.190	0.431	0.702	0.110	0.192	0.403

Table 5.3: Goodness-of-fit estimators for ordinal logistic models (top) and transition probabilities based on binary logistic models (bottom)

5.3.1 Actions on arrival on lower blinds

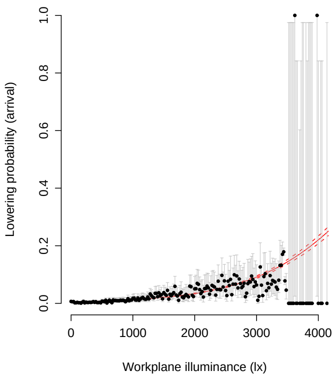
The best model accounting for the variation of these probabilities for lowering actions on arrival with a single predictor uses indoor horizontal illuminance E_{in} as the driving variable. Observed and fitted probabilities are shown in Figure 5.4(a), where we see that the probability of lowering a blind is well described by the fitted curve. We then consider a second variable, and observe that the initial lower unshaded fraction before action B_L induces the greatest increase in the predictive accuracy (Figure 5.4(c)). In other words the future position of the blind depends strongly on its previous position. Other variables do not bring any significant contribution if included as a third predictor.

For raising actions upon arrival we find that B_L is the most influential variable (Figure 5.4(b)), while a model with E_{in} fits only poorly. We also find that a model with both these variables offers a marginal but still significant improvement (Figure 5.4(d)). This suggests that if the occupants find their blind lowered on arrival they are more concerned with having an unobstructed view than by effective visual stimuli. Goodness-of-fit indicators (area under the ROC curve, Nagelkerke's R^2, Brier score and Somer's D_{xy}) are displayed in Table 5.3; the high values of AUC show that observed actions can be reliably reproduced while the values of R_N^2 are considered as good in the context of logistic regression models.

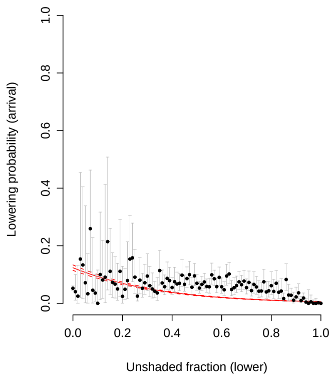
Note that the parameters displayed in Table 5.2 determine the final form of the model. Furthermore, the level plots of Figure 5.4 include the fitted regression surfaces which directly provide values for action probabilities given E_{in} and B_L.

5.3.2 Actions during presence and at departure on lower blinds

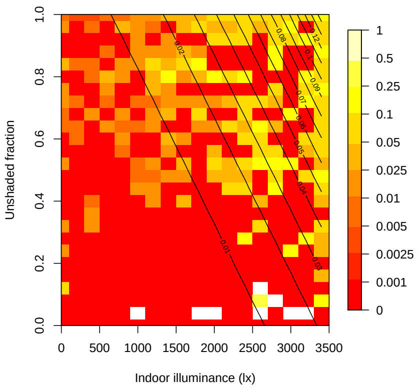
The variable selection process retains E_{in} and then B_L for lowering probability (Figure 5.4(e)), and B_L and then E_{in} for raising (Figure 5.4(f)); likewise on arrival. As expected from the observations in Table 5.1, the predicted probabilities are lower than on arrival, which confirms a specific behaviour in this situation. Goodness-of-fit indicators are also lower than on arrival (Table 5.3). Furthermore, some probability increase due to direct glare is evident when considering specific domains of ζ and α, which is discussed in Section 5.5.



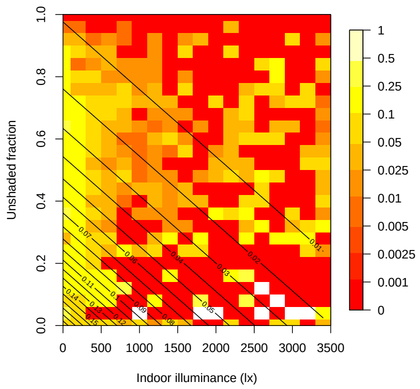
(a) Lowering actions on arrival versus indoor illuminance



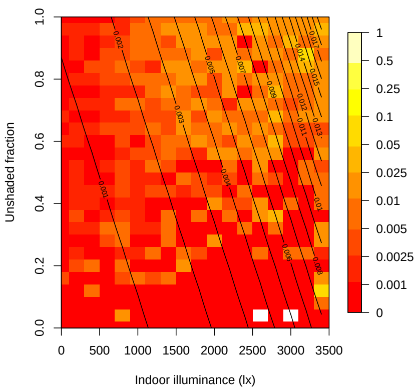
(b) Raising actions on arrival versus initial unshaded fraction



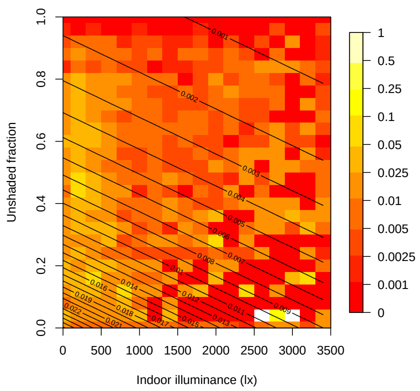
(c) Lowering actions on arrival versus indoor illuminance and initial unshaded fraction



(d) Raising actions on arrival versus indoor illuminance and initial unshaded fraction



(e) Lowering actions during presence versus indoor illuminance and initial unshaded fraction



(f) Raising actions during presence versus indoor illuminance and initial unshaded fraction

Figure 5.4: Level plots of observed action probabilites on lower blinds (given on a quasi-logarithmic scale), versus bins of indoor and outdoor temperature, with contour lines from the logistic regression surface

We have also examined the possibility of a purely seasonal effect on behaviour at departure, eg. whether occupants preventively lower their blinds when leaving during a heat wave to avoid heat gains during their absence. After examining actions with respect to both daily and monthly mean outdoor temperatures we did not notice any such behaviour.

Actions during presence thus do not differ fundamentally from actions on arrival, as the same variables influence these actions, but their frequency is lower. This may be seen by the roughly parallel contour lines from the regression surfaces in Figure 5.4 and from the regression parameters in Table 5.2.

5.3.3 Choice of lower blind position

We have also studied occupants' choices of unshaded fraction when performing an action. From this we observe (Table 5.1) that the general behaviour differs greatly in terms of whether a lowering or raising action is performed. We model the choice of unshaded fraction by first determining whether a full action (to fully raised or lowered) takes place. Once again we have used forward selection to identify key variables to infer a distribution for the probability of fully lowering and fully raising blinds, retaining $E_{\mathrm{gl,hor}}$ and B_L as predictors. The probability P_{full} of performing full lowering or raising actions can then be written as $\mathrm{logit}(P_{\mathrm{full}}) = a + b_{E_{\mathrm{out}}} E_{\mathrm{gl,hor}} + b_L B_L$, using the regression parameters given in Table 5.2. The observed proportion of full actions is shown in Figures 5.5(a)-5.5(b), together with contour lines from the fitted model.

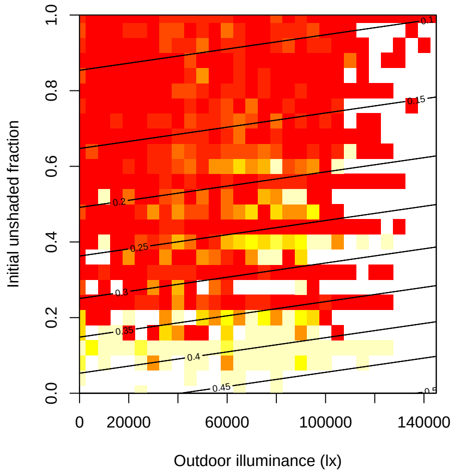
If a full action is not taken, a second sub-model must determine the shaded fraction from a relevant distribution. We studied observed lowering actions to a partial unshaded fraction, and observed that the increase in shading ΔB is well approximated by a Weibull distribution (see Section 3.4 for its definition), with scale parameter depending on the initial shaded fraction. Maximum likelihood estimation then yields:

$$f(\Delta B|B_{\mathrm{L,init}}) = \frac{\alpha}{\lambda(B_{\mathrm{L,init}})}(\frac{\Delta B}{\lambda(B_{\mathrm{L,init}})})^{\alpha-1}\exp(-(\frac{\Delta B}{\lambda(B_{\mathrm{L,init}})})^{\alpha}), \tag{5.3}$$

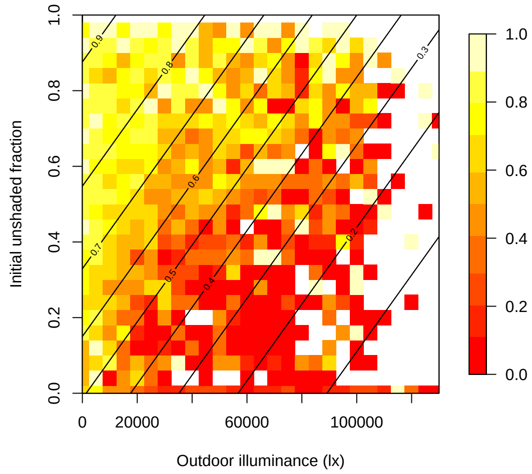
with shape $\alpha = 1.708$ and scale $\lambda(B_{\mathrm{L,init}}) = \exp(-2.294 + 1.522 \cdot B_{\mathrm{L,init}})$. The fitted distribution for ΔB is shown in Figure 5.5(c) for several values of $B_{\mathrm{L,init}}$. Figure 5.5(c) shows that occupants rarely choose slight or almost full lowering when setting their blinds until a non-total fraction.

But we have also found that the distribution of chosen fractions for partial raising actions does not significantly differ from a uniform distribution, which we use therefore in the final model. A refined treatment is anyway of less central importance, as partial raising actions are much less frequent than partial lowering actions (Table 5.1).

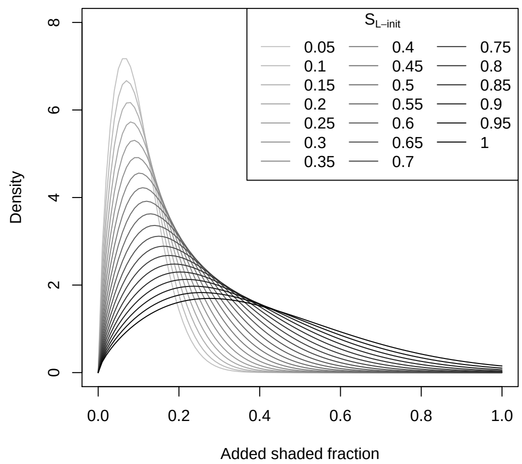
The approach of modelling first whether an action to fully raised or lowered fraction is performed and if this is not the case, following this by a second model to predict the chosen partial fraction has some drawbacks. It implies the use of two distinct stochastic models to predict unshaded fraction choices; it also neglects the significant correlation of the error terms in these models. However, the particularly high prevalence of full actions results in an extremely biased distribution of chosen fractions that would be difficult to model with a single process. Therefore, in spite of this additional complexity, the retained approach offers a sensible compromise between accuracy and accessibility for implementation.



(a) Probability of full lowering



(b) Probability of full raising



(c) Fitted distribution of added shaded fractions

Figure 5.5: Level plots of observed probabilites of full actions, versus bins of indoor illuminance and initial unshaded fraction, with contour lines from the logistic regression surface

5.3.4 Model for actions on upper blinds

We have performed similar analyses for the inference of a model for the upper blinds. We found that E_{in}, B_U and $E_{\mathrm{gl,hor}}$ were significant parameters for lowering and raising actions. Figure 5.6 shows observed and predicted action probabilities for domains of E_{in} and B_U.

For all these sub-models, the inclusion of $E_{\mathrm{gl,hor}}$ brings significant additional predictive value (Table 5.3), which was not the case for the lower blinds. This most likely relates to the particular purpose of the anidolic reflector to enhance internal daylight levels whilst $E_{\mathrm{gl,hor}}$ is low (and that of the blinds to prevent excess internal illumination whilst $E_{\mathrm{gl,hor}}$ is high).

The choice of upper shaded fraction is modelled similarly to those of lower blinds, with higher predicted probabilities of complete actions (Figures 5.6(e)-5.6(f)).

5.4 Predictive accuracy

Un coup de dés jamais n'abolira le hasard.
A roll of the dice will never abolish chance.
Poem (1897) of Stéphane Mallarmé (1842-1898)

5.4.1 Modelling variants

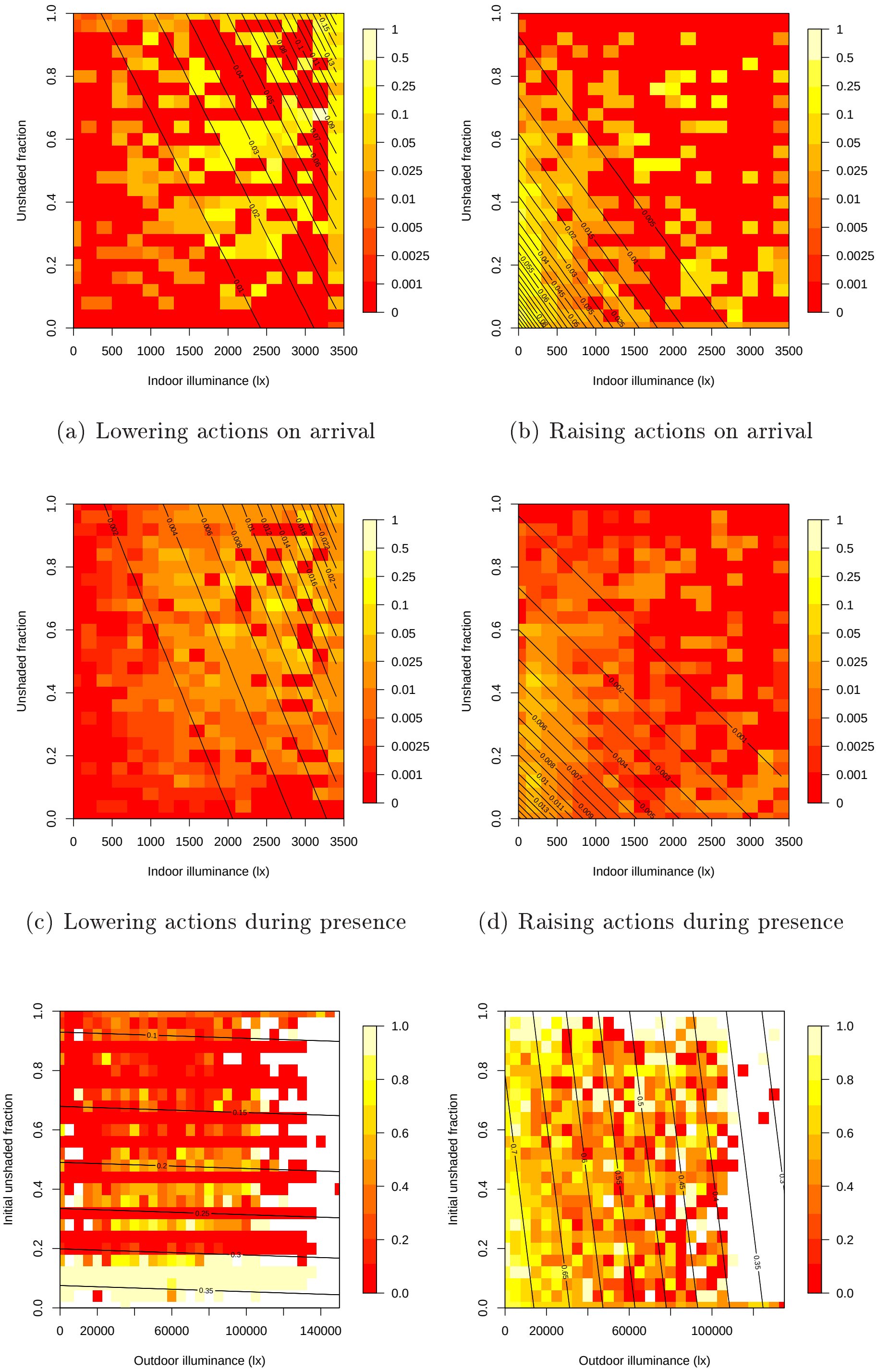
Based on these results, we have developed a simple algorithm for the simulation of blind usage, which is described in detail in Section 7.1 and graphically summarised in Figure 7.2. We consider four variants of decreasing complexity for this algorithm, in order to evaluate the accuracy induced by each of its elements:

- Model M1: We use the expressions inferred in Section 5.3 for P_{lower}, P_{raise}, $P_{\mathrm{full\,lower}}$, $P_{\mathrm{full\,raise}}$ and $f(\Delta B)$.
- Model M2: Same as M1, with $f(\Delta B)$ modelled as a uniform probability distribution.
- Model M3: Same as M2, with $P_{\mathrm{full\,lower}}$ and $P_{\mathrm{full\,raise}}$ treated as constants, with values from Table 5.1.
- Model M4: Same as M3, with P_{lower} and P_{raise} treated as constants, with values from Table 5.1.

5.4.2 Criteria

We describe here our procedure for rigorously evaluating the predictive power of our algorithm with its variants. We assess the predictive power of the models by checking four aspects:

- Overall activity: Is the predicted number of actions consistent with observations?
- Mean prediction: Does the model predict a consistent mean overall unshaded fraction throughout the simulation period?
- Distribution of shading: is the proportion of blinds fully lowered and fully raised acceptably reproduced?



(a) Lowering actions on arrival
(b) Raising actions on arrival
(c) Lowering actions during presence
(d) Raising actions during presence
(e) Probability of full lowering
(f) Probability of full raising

Figure 5.6: Level plots of observed action probabilites on upper blinds (top and middle) and probabilites of full actions (bottom), versus bins of indoor illuminance, initial unshaded fraction and outdoor illuminance, with contour lines from the logistic regression surface

Model	Lowering actions	Raising actions	Mean unshaded fraction	Prop. fully lowered	Prop. fully raised	Quartiles of error
Lower blinds						
Exact	5729	4838	84.6%	5.2%	67.1%	
M1	3919	3546	79.9%	2.3%	49.9%	-1.692 -0.946 -0.140
M2	3854	3859	79.2%	2.8%	52.5%	-1.859 -1.054 -0.220
M3	3980	3592	81.0%	2.7%	57.0%	-1.483 -0.809 -0.004
M4	3965	2436	55.7%	21.8%	37.9%	-4.615 -3.819 -2.944
Upper blinds						
Exact	6447	5556	67.6%	21.7%	56.4%	
M1	3799	2469	60.3%	9.8%	31.1%	-2.495 -1.300 0.413
M2	3774	2447	59.7%	10.0%	30.6%	-2.600 -1.506 0.457
M3	3706	2882	57.8%	21.5%	36.4%	-2.600 -1.461 0.074
M4	3863	2849	44.5%	44.2%	34.3%	-4.608 -2.923 -0.659

Table 5.4: Diagnostics for predictive accuracy: Mean simulated number of lowering actions, raising actions, average unshaded fraction, proportion of time fully lowered and fully raised and quartiles of the distribution of the error on aggregated unshaded fractions for lower (top) and upper (bottom) blinds

- Aggregated results: Is the predicted total unshaded fraction in the building consistent with observations?

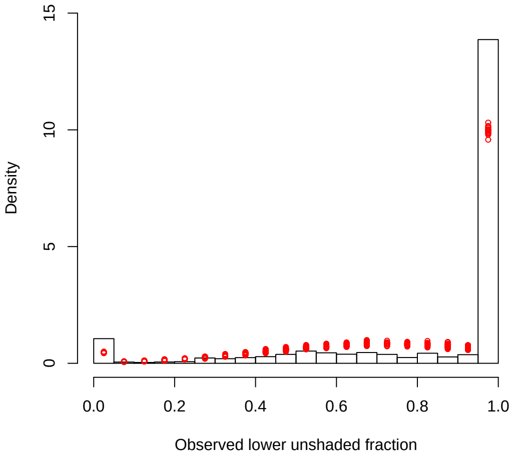
It is however not possible to perform a direct assessment of the associated discrimination as was the case in Chapter 4, as the predicted outcome is not binary for unshaded fractions.

In applying the above tests we have performed 20 repeated simulations using 5 minutes time steps for the whole period with available measurements and for the 12 measured offices, producing $20 \times 12 = 240$ sets of simulated lower and upper unshaded fractions $(B_{\mathrm{L,sim}}(t), B_{\mathrm{U,sim}}(t))$, to be compared with 12 sets of observed states $(B_{\mathrm{L,obs}}(t), B_{\mathrm{U,obs}}(t))$. This procedure was repeated for each of our models. Table 5.4 summarises the performance indicators, together with Figures 5.7 to 5.9.

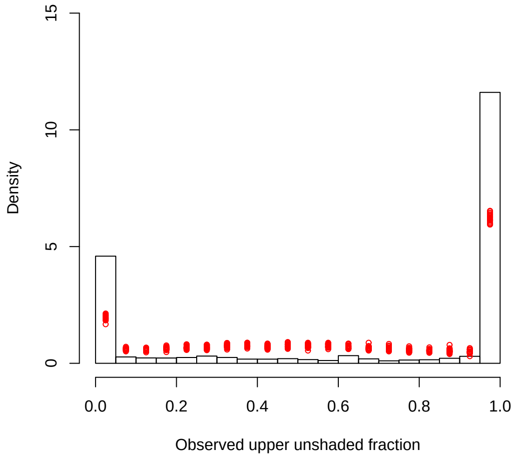
The total number of actions is underestimated for all models, indicating that these latter often neglect to predict an action that was performed. However missed actions will mostly result in comfortable daylight levels enabled by observed actions, which will decrease subsequent action probability (so that there is some in-built compensation). Nevertheless we observe that the magnitude of the estimation is coherent, and the AUC indices – measuring the proportion of correctly classified actions – range from from 78% to 89%.

The simulated mean unshaded fraction $(\langle B_{\mathrm{L,sim}}(t)\rangle, \langle B_{\mathrm{U,sim}}(t)\rangle)$ seems reliable, although it is slightly lower than observed. Our models based on visual stimuli (M1, M2 and M3) perform much better (Table 5.4) than the model with constant action probabilities (M4). From Figure 5.7(c) we can also see from the vertical clusters of results that this parameter does not vary strongly for repeated simulations of individual offices.

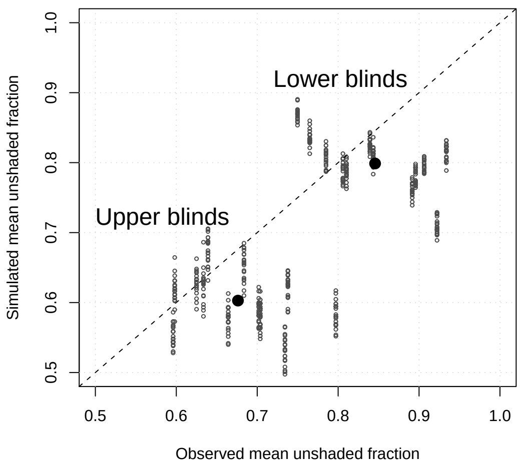
The simulated distribution of unshaded fractions for lower blinds also reproduces reliably the observed fractions, with low spread among simulation replicates. However, the predicted proportion of fully lowered and raised blinds is slightly too low (Table 5.4, 5$^{\text{th}}$ and 6$^{\text{th}}$ columns). The obtained shading distribution is less satisfactory for upper blinds. However this is not problematic in this particular case, since the specific design of the anidolic windows is such that unshaded fractions between 0 and 0.6 have little impact on solar gains (the zenith angle from the centre of the vertical plane of the window to the base of the blinds remains small). In general, the model M4 with constant probabilities performs poorly, which underlines the added value of visual stimuli in a model for actions



(a) Lower blinds: histograms of observed (black) and simulated (red) unshaded fractions



(b) Upper blinds: histograms of observed (black) and simulated (red) unshaded fractions



(c) Observed versus ten simulated mean unshaded fractions for all offices

Figure 5.7: Graphical diagnostics for validation using the model M1

on blinds.

We extend empirical comparisons to include simulation to the scale of the whole building and check whether the models are able to reproduce the total observed shading levels for all offices. Based on individual simulations of blind states in the 12 offices, we compute the total unshaded fractions $B_{\mathrm{sim}}(t) = (\sum_{k=1}^{12} B_k(t))/12$ for each simulation. Averaging over the 20 sets of $B_{\mathrm{sim}}(t)$ we deduce the mean predicted unshaded fraction at each time step $\bar{B}_{\mathrm{sim}}(t)$ (Figure 5.8(c)). Comparing with the observed total unshaded fraction $B_{\mathrm{obs}}(t)$, we may compute the mean error $\varepsilon(t) = \bar{B}_{\mathrm{sim}}(t) - B_{\mathrm{obs}}(t)$ at each time step (Figures 5.8(a)-5.8(b)).

We show in Table 5.4 the quartiles of this error for each model. From this we observe that the full model (M1) and a partial model (M3) offer the best performance on aggregated results, with smallest error magnitude. Furthermore, comparisons of measured and predicted unshaded fractions (Figures 5.9(a)-5.9(b)) indicate that these models reproduce acceptably well the temporal variation of blinds' position[3].

In conclusion, the models including visual stimuli in action probabilities (M1, M2 and M3) offer more robust predictions and their use is essential for integration in building simulation. The performance of the most refined models (M1 and M2) is however similar compared to the more basic model M3.

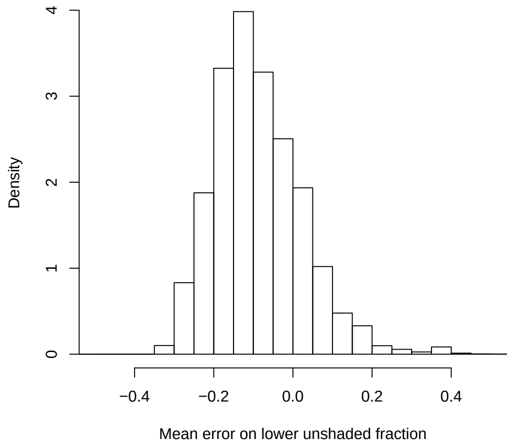
5.5 Discussion

Like most previous surveys, our measurements come from a single configuration of blinds, which potentially limits the generality of our results. There is a great diversity in the available types of shading devices, and differences in their use are evident (eg. occupants may interact differently with curtains, blinds with blades, Venetian blinds or shutters). Furthermore, our data are also currently restricted to an office type environment; in which the importance of other factors such as privacy may not be as high as they are in the home. The type of control may also play a role, as occupants of our survey may easily adjust their blinds simply by pressing a button. We cannot exclude that the ease of use of shading devices influences the probabilities of action and the chosen unshaded fractions. It is also possible that measurements in other climatic contexts would lead to dissimilar results, for instance when actions on air conditioning systems are favoured in lieu of shading devices or when the sunpath is very different.

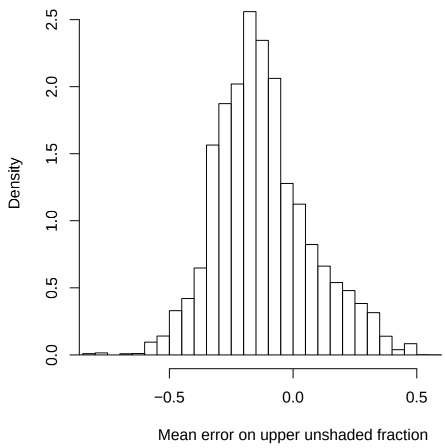
These issues bring out the necessity to develop a general method supported by a large dataset including measurement for diverse configurations of blind and building context. Nevertheless, our approach based on local stimuli (indoor horizontal illuminance) offers promise for extension to other configurations of shading device and building context. Indeed we may expect to find the same driving variables, but with different action probabilities depending on shading device accessibility. The most building-specific feature is expected to be the choice of shaded fraction.

Our transition probabilities do not include outdoor radiation parameters such as $I_{\mathrm{gl,hor}}$, $I_{\mathrm{diff,hor}}$ or I_{beam}. Our statistical analyses showed that the relevant visual stimuli are already included with E_{in} and $E_{\mathrm{gl,hor}}$. Furthermore, even if they were found to be more influential, their use would be problematic as their influence on occupants' actions varies with the glazed surface and the current shaded fraction, which would make their prediction

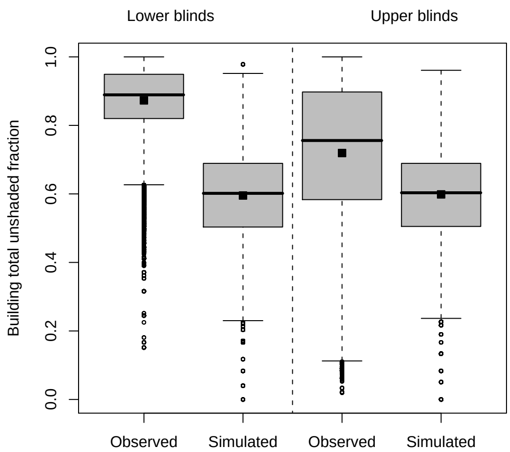
[3] This refers to the period from 27 January 2005 to 14 January 2006, which offers representative climatic conditions and uninterrupted measurements.



(a) Histogram of the error on total lower unshaded fractions

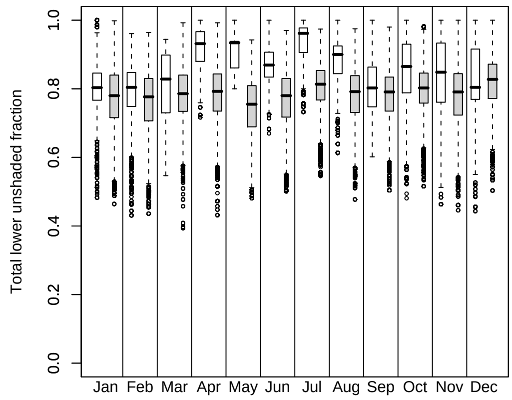


(b) Histogram of the error on total upper unshaded fractions

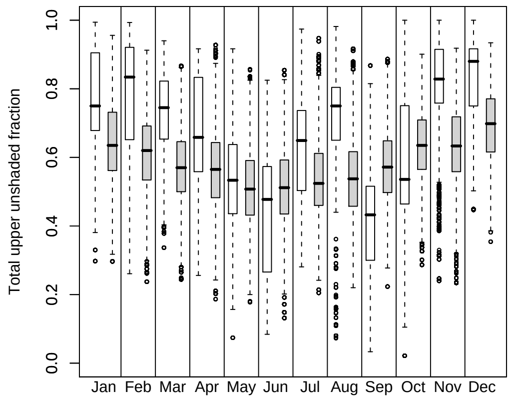


(c) Boxplots of observed and simulated total unshaded fractions

Figure 5.8: Simulation results of the model M1 at the scale of the whole building



(a) Observed and mean simulated total lower unshaded fractions on a one year period



(b) Observed and mean simulated total upper unshaded fractions on a one year period

Figure 5.9: Monthly box-and-whisker plots of observed and mean simulated total unshaded fractions on a period of a year using the model M1

purely specific to our building. The use of indoor visual stimuli is thus more coherent, provided that a daylight model is coupled with simulations. Although horizontal workplane illuminance is the surveyed environmental variable that best explains actions on shading devices, there may well exist better predictors, such as the CIE glare index (CGI) [99, 100] or the CIE unified glare rating system (UGR) [101]:

$$\mathrm{CGI} = 8\log_{10} 2\frac{1+(E_d/500)}{E_d+E_i}\sum_{i=1}^{n}\frac{L_s^2\omega_s}{P^2}, \quad \mathrm{UGR} = 8\log_{10}\frac{0.25}{L_b}\sum_{i=1}^{n}\frac{L_s^2\omega_s}{P^2}, \tag{5.4}$$

where E_d and E_i are the direct and indirect vertical illuminances at the eye, L_s the luminance of the glare source, ω_s the solid angle subtended by the source, L_b the general field of luminance controlling the adaptation levels of the observer's eye and P Guth's position index (depending on the observer's line of sight and the azimuth and elevation of the source). Another interesting indicator is daylight glare probability (DGP), proposed by Wienold and Christoffersen [102] who suggested that it better relates with glare sensation. It is defined as

$$\mathrm{DGP} = c_1 E_v + c_2 \log\left(1 + \sum_i \frac{L_{s,i}^2\omega_{s,i}}{E_v^{c_4} P_i^2}\right) + c_3, \tag{5.5}$$

with $c_1 = 5.87 \cdot 10^{-5}$, $c_2 = 9.18 \cdot 10^{-2}$, $c_3 = 0.16$ and $c_4 = 1.87$. Furthermore, its definition as a probability is interesting in relationship with action probabilities. However, its applicability to our database is undermined by the fact that variables such as L_s and E_v cannot be deduced from our measurements, even if we have a mean to deduce the solid angle ω_s. Furthermore, it may not be directly applicable to our situation, as our occupants sit perpendicular to the window and thus glare occurs out of the main field of vision.

The sun position (ζ, α) was not found to be a significant addition as raw variables to models including local visual stimuli and initial blind position. Figure 5.10 shows the

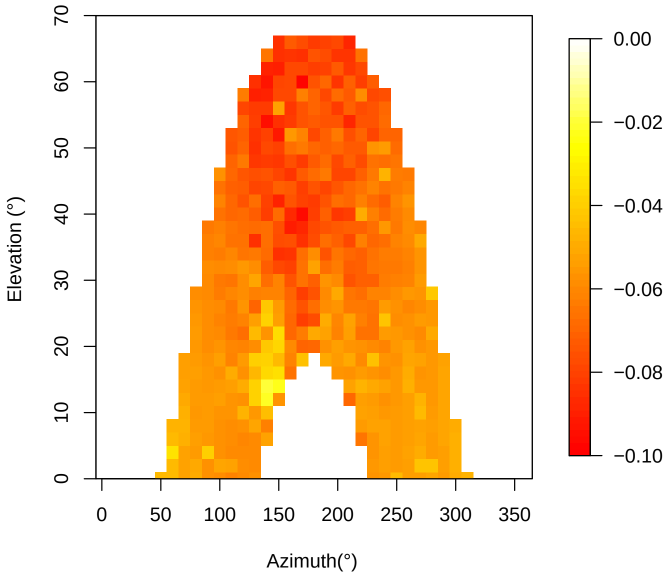


Figure 5.10: Level plot of deviance residuals of the model for lowering actions during presence including E_{in} and B_L, with respect to sun position

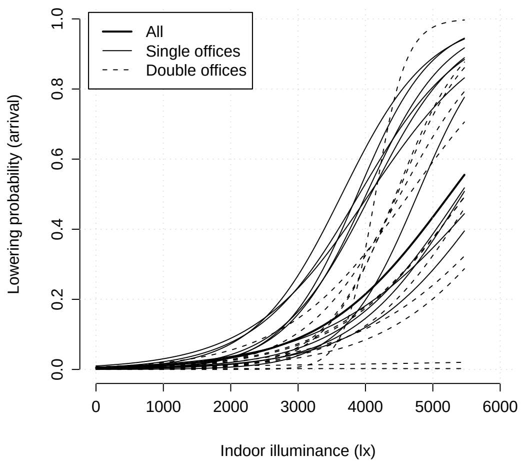
deviance residuals (see Section 3.2.5) of the model for lowering actions during presence including E_{in} and B_L. A domain centered on $\zeta = 15°$, $\alpha = 140°$ shows a strongly increased observed action probability not treated by the model, which is likely to correspond to the sun positions involving direct glare, based on the configuration of the offices. The present formulation of our model does not include the treatment of actions in response to direct glare risk; however a modification of the model to include this stimulus would be straightforward.

Finally, our models do not include thermal stimuli, as visual variables alone explain observed variations in actions on blinds. However, these parameters could play a direct role in buildings which utilise daylight less efficiently; where occupants would need to lower blinds for thermal reasons rather than in response to perceived glare.

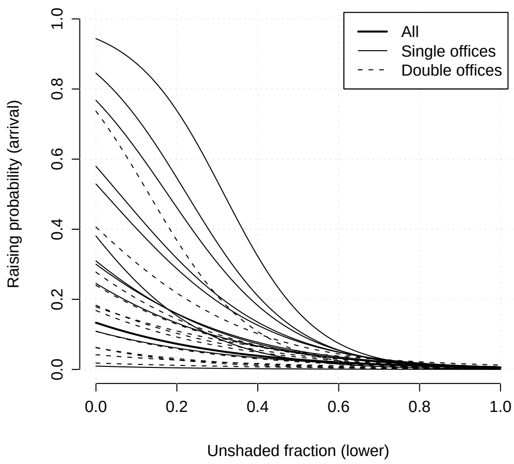
5.6 Individuals' behaviours

The models developed above were derived from data relating to the whole set of occupants and for the entire surveyed period. We examine here variations in behaviour among the surveyed occupants and provide a method to account for observed behavioural diversity amongst them.

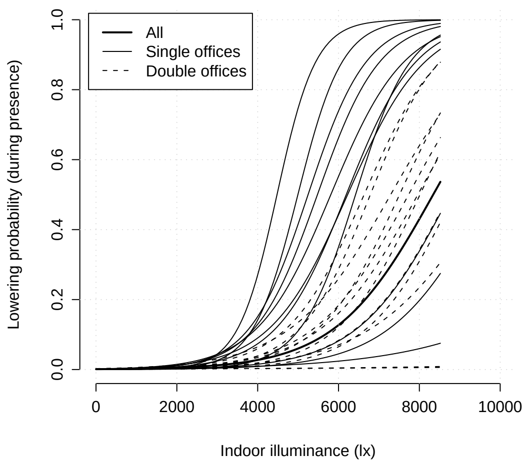
Table 5.5 shows for each individual or combination of occupants (in double offices) the regression parameters for univariate probabilities of actions on lower blinds, based on the most influential predictor only (E_{in} for lowering and B_L for raising actions). The obtained curves are shown in Figure 5.11, where it is also evident that for the majority of occupants the slopes b of the action probabilities are reasonably similar; the differences arising mainly in the intercepts a.



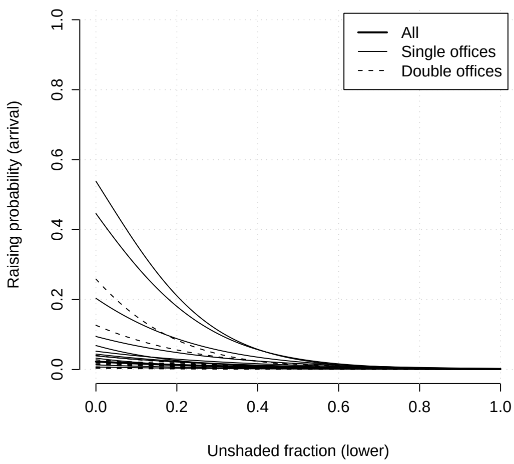
(a) Lowering on arrival transition probabilities



(b) Raising on arrival transition probabilities



(c) Lowering during presence transition probabilities



(d) Raising during presence transition probabilities

Figure 5.11: Occupant specific action probabilities based on their most influential variables

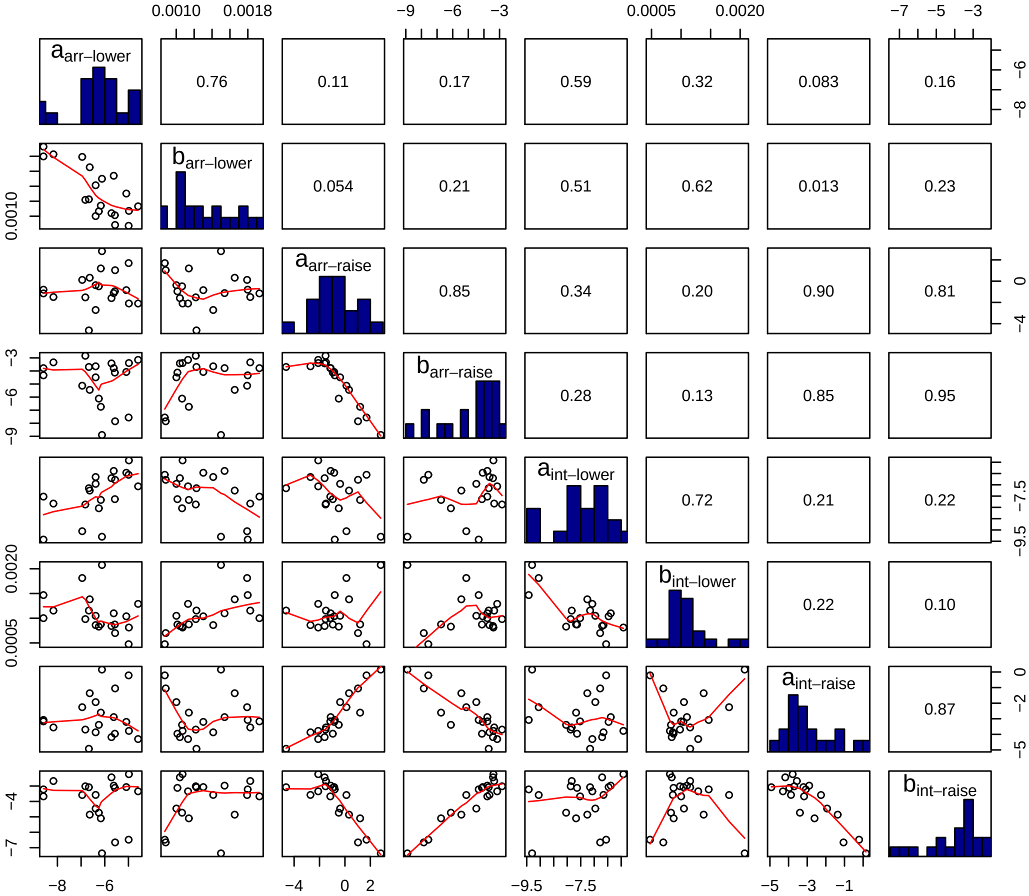


Figure 5.12: Bivariate plots between all individual regression parameters, with local polynomial regression, correlations and histograms

Pers	$\mathbf{P}_{\mathbf{lower,arr}}$		$\mathbf{P}_{\mathbf{raise,arr}}$		$\mathbf{P}_{\mathbf{lower,int}}$		$\mathbf{P}_{\mathbf{raise,int}}$	
Ref.	a	$b \cdot 10^3$	a	$b \cdot 10^3$	a	$b \cdot 10^3$	a	$b \cdot 10^3$
001-28-X	-6.39 ± 0.42	1.41 ± 0.14	-2.71 ± 0.19	-3.64 ± 0.58	-6.66 ± 0.16	0.86 ± 0.07	-3.87 ± 0.08	-3.58 ± 0.26
001-17-30	-8.17 ± 0.81	1.83 ± 0.26	-1.49 ± 0.24	-3.34 ± 0.37	-7.83 ± 0.23	1.15 ± 0.08	-3.56 ± 0.12	-2.68 ± 0.17
001-17-35	-4.98 ± 0.48	1.07 ± 0.20	-2.10 ± 0.38	-3.39 ± 0.84	-5.92 ± 0.25	0.81 ± 0.12	-3.78 ± 0.20	-2.23 ± 0.37
002-21-24	-6.39 ± 0.75	1.00 ± 0.27	-0.38 ± 0.45	-4.49 ± 0.61	-6.96 ± 0.27	1.05 ± 0.11	-1.93 ± 0.26	-4.46 ± 0.37
002-16-25*	18.07 ± 11.60	4.36 ± 3.40	-3.13 ± 0.56	-2.36 ± 0.80	-8.26 ± 0.42	0.93 ± 0.17	-4.87 ± 0.31	-2.32 ± 0.41
002-01-16	-6.82 ± 1.69	1.22 ± 0.54	-1.53 ± 0.45	-2.84 ± 0.75	-7.87 ± 0.77	0.98 ± 0.25	-3.69 ± 0.31	-3.02 ± 0.55
002-11-16	-8.59 ± 1.32	1.93 ± 0.41	-1.15 ± 0.40	-3.78 ± 0.56	-7.52 ± 0.28	1.00 ± 0.11	-3.17 ± 0.28	-3.67 ± 0.37
003-07	-5.09 ± 0.14	1.30 ± 0.06	-0.85 ± 0.14	-4.08 ± 0.24	-6.47 ± 0.10	1.04 ± 0.05	-3.20 ± 0.08	-3.07 ± 0.15
004-26-36	-5.72 ± 0.17	1.04 ± 0.07	-1.60 ± 0.14	-3.42 ± 0.21	-6.71 ± 0.10	0.84 ± 0.05	-4.18 ± 0.09	-2.44 ± 0.14
101-22	-6.67 ± 0.31	1.23 ± 0.14	-4.63 ± 0.32	-3.69 ± 0.91	-7.15 ± 0.14	1.16 ± 0.08	-4.92 ± 0.10	-3.08 ± 0.27
103-03	-5.62 ± 0.20	1.54 ± 0.08	-1.12 ± 0.18	-3.78 ± 0.30	-6.39 ± 0.09	1.10 ± 0.04	-2.89 ± 0.08	-2.98 ± 0.12
103-42	-6.12 ± 1.24	1.50 ± 0.41	2.82 ± 1.99	-8.90 ± 3.15	-9.30 ± 1.15	2.08 ± 0.39	0.15 ± 0.76	-7.37 ± 1.27
103-29	-4.59 ± 0.65	1.13 ± 0.30	-2.10 ± 0.43	-3.15 ± 1.19	-7.07 ± 0.53	1.29 ± 0.23	-4.30 ± 0.31	-3.34 ± 0.96
104-19	-6.64 ± 0.21	1.65 ± 0.08	0.32 ± 0.19	-5.45 ± 0.27	-7.25 ± 0.11	1.38 ± 0.05	-1.36 ± 0.11	-4.86 ± 0.15
105-38	-8.59 ± 0.61	1.80 ± 0.19	-0.80 ± 0.28	-4.34 ± 0.34	-9.43 ± 0.38	1.47 ± 0.12	-3.07 ± 0.14	-3.22 ± 0.19
106-06-39	-5.57 ± 0.20	1.01 ± 0.08	-0.96 ± 0.20	-4.13 ± 0.28	-7.64 ± 0.16	0.87 ± 0.07	-3.96 ± 0.12	-3.12 ± 0.18
201-31	-6.25 ± 0.25	1.07 ± 0.10	-0.49 ± 0.30	-6.12 ± 0.44	-8.04 ± 0.16	0.83 ± 0.06	-3.39 ± 0.24	-4.73 ± 0.30
202-05-33	-5.56 ± 0.41	0.88 ± 0.16	1.03 ± 0.80	-7.85 ± 1.40	-6.78 ± 0.21	0.70 ± 0.09	-1.05 ± 0.33	-6.66 ± 0.57
203-09	-4.99 ± 0.27	0.87 ± 0.11	1.70 ± 0.68	-7.56 ± 0.82	-6.56 ± 0.18	0.47 ± 0.09	-0.22 ± 0.27	-6.48 ± 0.33
203-08-14*	-4.91 ± 0.42	0.19 ± 0.25	-3.98 ± 0.70	-2.23 ± 1.27	-6.53 ± 0.19	0.17 ± 0.12	-5.18 ± 0.25	-2.13 ± 0.46
203-12-40*	-5.94 ± 1.74	-0.01 ± 1.13	-2.70 ± 0.60	-4.36 ± 1.86	-6.83 ± 0.56	0.25 ± 0.31	-5.60 ± 0.54	-3.59 ± 1.48
204-18	-6.95 ± 0.60	1.79 ± 0.20	0.12 ± 0.52	-5.13 ± 0.73	-9.06 ± 0.50	1.81 ± 0.17	-2.26 ± 0.25	-3.58 ± 0.34
204-10	-6.17 ± 0.30	1.14 ± 0.10	1.20 ± 0.28	-6.74 ± 0.45	-7.67 ± 0.20	0.88 ± 0.07	-2.61 ± 0.25	-5.10 ± 0.39

Table 5.5: Occupant specific parameters for action probabilities on lower blinds including a single variable. Asterisks are used to identify occupants for which at least one regression parameter was not statistically significant.

Figure 5.12 shows the main patterns linking these regression parameters. For all cases, intercepts and slopes are strongly correlated; it is thus meaningful to summarise the variability between occupants by the characteristic illuminance $E_{50} = -a/b$ for which the probability of action is equal to 0.5. This allows an indirect measure of the notion of "activity" or "passivity" of individual occupants, introduced by Reinhart [78, 79]; those displaying a low value of E_{50} being more likely to perform lowering actions at less extreme illuminance levels.

It is particulary noticeable in Figures 5.11(a) and 5.11(c) that occupants of single offices display higher values of E_{50} and thus act on their blinds with higher probability. The strong correlations between all regression parameters for arrival and during presence (Figure 5.12) are also interesting: occupants individually display a coherent level of adaptive activity in both these situations, action probability being higher on arrival.

5.7 Conclusion

We have developed a model for the prediction of actions on shading devices based on a long term survey. It is formulated as transition probabilities, based on a sufficient set of driving variables. This model represents a clear improvement compared to currently used deterministic methods. We have also rigorously tested this model against measured data with satisfactory results.

The model is based on general assumptions, for application to a wide class of buildings. Although our data relate to a very specific building design, group of occupants and shading system, we have nevertheless found conclusively that the driving variables for actions are local stimuli on the workplane, which directly links visual comfort, visual variables and actions. The form of the model should thus be readily adaptable to other situations, but to be able to model the range of situations with confidence considerably more data is required with which to derive the relevant model calibration parameters. The model may also be integrated with any dynamic building thermal simulation program which includes predictions of indoor illuminance. Finally, we have also studied individuals' diversity in the use of shading devices and suggested how this diversity may be accounted for in the model.

Chapter 6

Environmental comfort

Il faut un minimum de confort pour pratiquer la vertu.
Virtue requires a minimum of comfort.
Thomas Aquinas (ca. 1225-1274), Opera omnia

Environmental comfort and indoor environment quality more generally are important factors in building performance simulation, as the final purpose of a building is to ensure the comfort of its occupants. In this chapter, we propose a global approach to better understand and predict occupant comfort, and account for its close relationship with the adaptive actions treated in Chapters 4 and 5.

This chapter begins with a brief review of the relevant research in the fields of thermal and visual comfort (Section 6.1). Based on preliminary observations and comparisons with current accepted adaptive models for thermal comfort, we underline the need for further research (Section 6.2). From our data, we go on to compare different approaches for the prediction of thermal comfort (Section 6.3), including an explicit distribution of thermal sensation and an analytical probability of thermal comfort, followed by an analysis of the influence of adaptive actions on comfort temperature (Section 6.4)[1]. A similar approach is then applied to the prediction of visual sensation and comfort (Section 6.5). We conclude by proposing a general formulation of control action probability based on the perception of comfort and action inertia (Section 6.6), where occupant comfort is put in its appropriate place as the central underlying concept in occupants' behaviour and adaptation. We go on to discuss how environmental as well as personal interactions influence human comfort, which in turn may influence subsequent control actions. We finally discuss future perspectives for research in environmental comfort (Section 6.7).

6.1 State of the art

We present in this section the key definitions related to thermal (Section 6.1.1) and visual (Section 6.1.2) comfort and discuss key outcomes from research that has been conducted until now.

[1]The methodology presented in Section 6.4 has been presented at the 10$^{\text{th}}$ International Building Performance Simulation Association Conference [103], at the CISBAT Conference 2007 [104] and published in the journal *Building and Environment* [74].

6.1.1 Thermal comfort

> Ce qui a été cru par tous, et toujours, et partout,
> a toutes les chances d'être faux.
> *That which has always been accepted by everyone,*
> *everywhere, is almost certain to be false.*
> Paul Valéry (1871-1945), Tel Quel (1943)

Sensation, comfort and acceptability

Thermal comfort is defined by ASHRAE as *that state of mind which expresses satisfaction with the thermal environment* [105]. As thermal comfort is based on perception it must be measured subjectively. Most studies and standards use the seven-point ASHRAE or the Bedford scale [35, 106] (Table 6.1) to quantify thermal sensation, from which our questionnaire (Table 2.3) is inspired.

Rating	ASHRAE scale	Bedford scale	Satisfaction
−3	Cold	Much too cool	Not satisfied
−2	Cool	Too cool	(too cold)
−1	Slightly cool	Comfortably cool	
0	Neutral	Comfortable	Satisfied
+1	Slightly warm	Comfortably warm	
+2	Warm	Too warm	Not satisfied
+3	Hot	Much too warm	(too hot)

Table 6.1: Thermal sensation and satisfaction scales

For both scales, it is assumed that a person reporting one of the three central categories (−1 to +1) is in a state of thermal comfort. Previous studies generally observed the close agreement between the notions of comfort and acceptability. For instance, Berglund [107] found that subjects' thermal comfort votes rated as comfortable or slightly uncomfortable were perceived as acceptable. More recently, based on an experiment with stationary conditions in a climate chamber, Zhang and Zhao [108] concluded that the notions of thermal sensation, acceptability and comfort seem to be equivalent for uniform environments, while in the non-uniform case a limited deviation regarding overall thermal sensation was noticed (explained by non-uniformity of thermal sensation). Similar discrepancies exist in the case of transient environments.

Several studies have shown that the central category may not necessarily always be the preferred subjective thermal sensation. For example, based on a simultaneous survey of thermal sensation and preference, Humphreys and Nicol observed that in the case of warm indoor temperatures, occupants' preferred sensation tended to be slightly warmer than neutral [109].

Fanger's stationary heat balance model

The model of Fanger [106], subsequently integrated to the ISO 7730 Standard [35], proposes a model based on a stationary human thermal heat balance to compute the predicted mean vote (PMV) of a population of occupants – the mean thermal sensation in a given environment – based on the ASHRAE sensation scale (Table 6.1).

Variable	Symbol, unit	Domain of validity
Air temperature	T_a [K], θ_a [°C]	10 to 30°C
Mean radiant temperature	T_{mr} [K], θ_{mr} [°C]	10 to 40°C
Relative air speed	v [m/s]	0 to 1 [m/s]
Partial vapour pressure	p [Pa]	0 to 2700 [Pa]
Specific activity	m [W/m^2]	46 to 230 [W/m^2] (0.8 to 4 met)
Specific work	w [W/m^2]	
Clothing level	I_{cl} [m^{2}K/W]	0 to 0.310 [m^{2}K/W] (0 to 2 clo)

Table 6.2: Parameters for the Fanger equation

Based on extensive studies of subjects placed in climate chambers, Fanger derived an equation for the PMV using the following heat balance equation with parameters given in Table 6.2,

$$\begin{aligned} PMV &= (0.303\exp(-0.036m)+0.028)\cdot \\ &\cdot\Big(m-w-0.00305(5733-699(m-w)-p)-0.42(m-w-58.15) \\ &-1.7\cdot 10^{-5}m(5867-p)-0.0014m(307-T_a) \\ &-3.96\cdot 10^{-8}f(T_{\mathrm{cl}}^4-T_{\mathrm{mr}}^4)-fh(T_{\mathrm{cl}}-T_a)\Big), \end{aligned} \tag{6.1}$$

in which f is the fraction of clothed surface, h the convective heat transfer coefficient and T_{cl} the surface temperature of clothing, which are given by

$$f=\begin{cases} 1.00+1.290\cdot I_{\mathrm{cl}} & (\text{for } I_{\mathrm{cl}}<0.078\ \mathrm{m^2K/W}) \\ 1.05+0.645\cdot I_{\mathrm{cl}} & (\text{for } I_{\mathrm{cl}}>0.078\ \mathrm{m^2K/W}) \end{cases}$$

$$h = \max(2.38(T_{\mathrm{cl}}-T_a)^{1/4}, 12.06\cdot v), \tag{6.2}$$

$$T_{\mathrm{cl}} = 308.9-0.028(m-w)-I_{\mathrm{cl}}\cdot\big(3.96\cdot 10^{-8}f(T_{\mathrm{cl}}^4-T_{\mathrm{mr}}^4)+fh(T_{\mathrm{cl}}-T_a)\big). \tag{6.3}$$

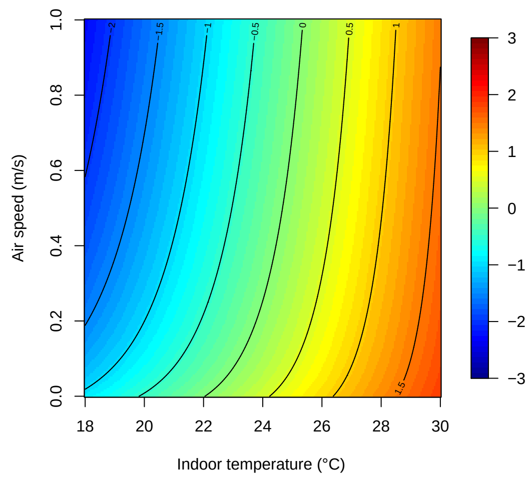
Equation (6.1) is obtained through multiplying a coefficient for change in metabolic rate: $(0.303\exp(-0.036m)+0.028)$, by the difference between heat production $(m-w)$ and contributions to heat losses (appearing sequentially in Equation (6.1): vapour diffusion, sweat evaporation, latent respiration, dry respiration, radiation and convection).

The partial vapour pressure can be obtained from the relative humidity ϕ through $p=p_{\mathrm{sat}}(T)\phi$, and the saturation vapour pressure p_{sat} can be deduced through the Goff-Gratch equation [110],

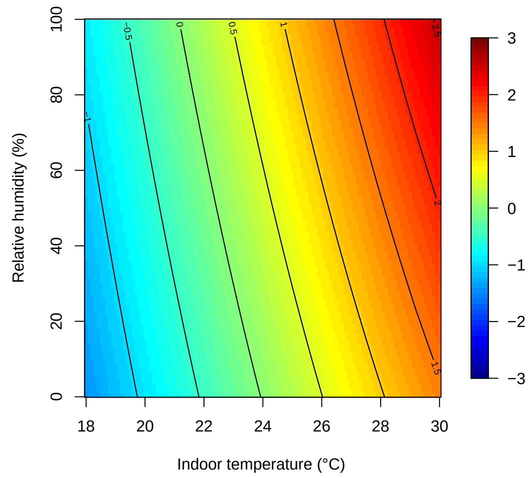
$$\begin{aligned} \log_{10}(p_{\mathrm{sat}}) &= -7.90298\cdot(\frac{373.16}{T}-1)+5.02808\cdot\log_{10}(373.16/T)-1.3816\cdot 10^{-7} \\ &\cdot(10^{11.344\cdot(1-\frac{T}{373.16})}-1)+8.1328\cdot 10^{-3}\cdot(10^{-3.49149\cdot(\frac{373.16}{T}-1)}-1)+\log_{10}(1013.246) \end{aligned} \tag{6.4}$$

This yields an implicit formulation of the PMV, therefore Equations 6.1 to 6.3 must be solved iteratively with a computer[2]. We show in Figure 6.1 calculations of the PMV for several domains of its driving variables.

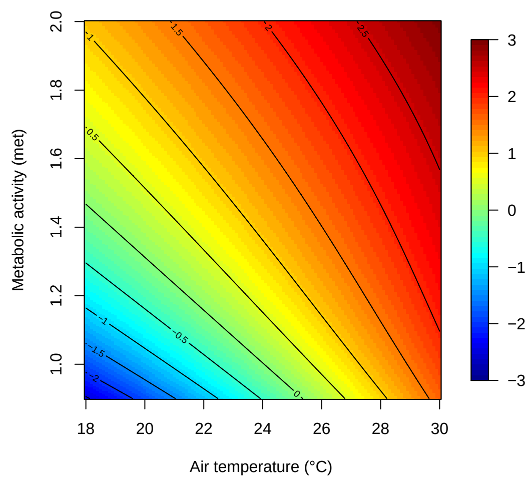
[2]However, a simplified approach proposed by Sherman [111] allows a reasonably precise computation of PMV without iteration, in which the PMV is a linear function of the air temperature and mean radiant temperature and a quadratic of the dew point.



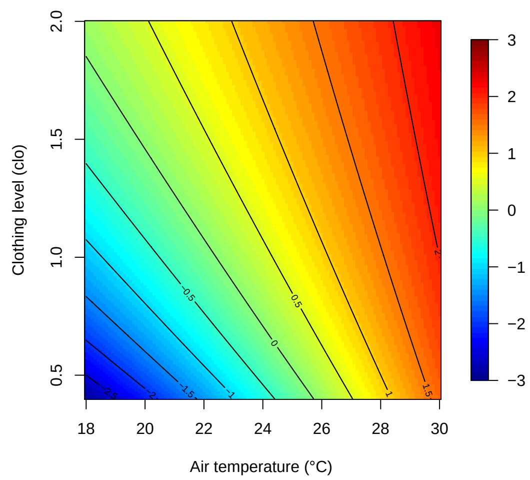
(a) PMV as a function of air temperature and air speed, with fixed ϕ, m, I_{cl}



(b) PMV as a function of air temperature and relative humidity, with fixed v_a, m, I_{cl}



(c) PMV as a function of air temperature and metabolic activity, with fixed v_a, ϕ, I_{cl}



(d) PMV as a function of air temperature and clothing level, with fixed v_a, m, ϕ

Figure 6.1: Fanger's PMV predictions, assuming $\theta_{\text{mr}} = \theta_a$, $w = 0$ W/m^2, $I_{\text{chair}} = 0.13$ clo, and when fixed, $\phi = 50\%$, $v_a = 0$ m/s, $m = 1.2$ met, $R = 0.7$ clo,

Optimal conditions correspond thus to a zero PMV. A related parameter for assessing global thermal comfort is the predicted percentage of dissatisfied (PPD) in a population, which can be predicted from the PMV through the following empirical relationship:

$$PPD = 1 - 0.95 \cdot \exp(-0.003353 \cdot PMV^4 - 0.2179 \cdot PMV^2). \tag{6.5}$$

According to which we have at best 5% of dissatisfied people, corresponding to $PMV = 0$, which expresses the impossibility of creating optimal conditions satisfying a whole set of occupants, but that it is possible to create an environment where the percentage of satisfied is maximal.

Limitations of Fanger's model

The model of Fanger reproduces generally well observed thermal sensation and satisfaction in controlled steady state environments such as air-conditioned buildings. However, its predictions are known to be generally poor in naturally-ventilated buildings, where the environment is more variable. Several causes for this failure to provide accurate predictions in such environments have been put forward (for instance by Newsham [112], Olesen and Parsons [113], Humphreys and Nicol [114] and van Hoof [115]):

- Fanger's model is based on a simplified calculation of the heat balance of the human body, which does not account for any inertia in heat storage. This can lead to significant errors in transient environments.

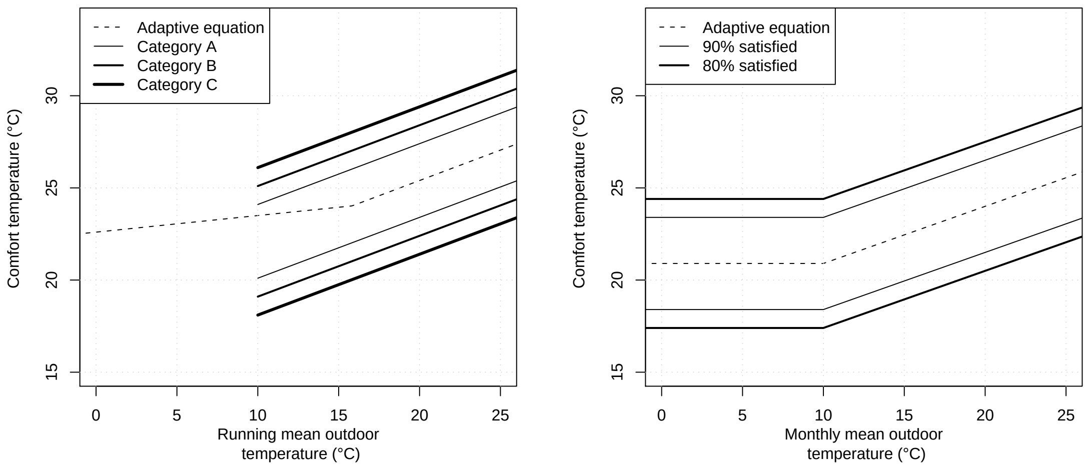
- The prediction of thermal sensation away from neutrality is based upon the principle of thermal load (Equation (6.1)), which is criticised.

- The model ignores the processes of acclimatisation; that is, occupants' thermal expectations may evolve in relation to the climatic context and their thermal experience.

- The model does not account for occupants' ability to adjust their environment to restore their thermal comfort through the use of buildings controls (windows, blinds, fans) or their personal characteristics (clothing). Furthermore, the mere knowledge that these adaptive opportunities are available may enable occupants to feel comfortable for a wider range of temperatures.

In particular, numerous field studies have underlined the positive and crucial impact of adaptive opportunities on occupant comfort. For instance, Oseland [116] and Leaman and Bordass [117] found that occupants' perceived ability to control their environment significantly impacts their satisfaction with that environment, and that occupants were more forgiving of discomforting influences in buildings that provided good opportunities for occupant control. Quantitatively, Brager et al. [118] observed a 1.5°C difference in the neutral temperature between occupants with high and low degree of control over windows. More recently, Karjalainen noticed that perceived control was higher and room temperature satisfaction better in homes than offices, where adaptive actions are less constrained [119].

This decisive impact of adaptive opportunities on the sensation of comfort – physical, physiological and psychological – motivated the development of other methods for predicting thermal comfort. Called adaptive comfort models these are presented below.



(a) CEN standard EN15251

(b) ASHRAE 55-2004 standard

Figure 6.2: Adaptive comfort temperature according to international standards

Adaptive comfort models

Adaptive comfort models attempt to account for observed variations in comfort temperature induced by occupants' ability to adapt their environment, in order to overcome the limitations of Fanger's model. These adaptive opportunities induce a change of paradigm in the understanding of thermal sensation, summarised by Nicol and Humphreys [120] in 1973, who stated that *"subjective warmth should be seen as an active link in the control system, and not as a mere passive response to the thermal environment"*.

Adaptive comfort models typically predict a *neutral* or *comfort temperature*, defined as "the operative temperature at which either the average person will be thermally neutral or at which the largest proportion of a group of people will be comfortable" [121]. Until now, this comfort temperature has, based on its observed seasonal dependence, been related to some running mean value of outdoor temperature.

CEN standard. The recent CEN standard [122] uses the exponentially weighted running mean of the daily mean air temperature as a predictor, as follows:

$$\theta_{\mathrm{rm}} = (1-\alpha)\theta_{\mathrm{od-1}} + \alpha\theta_{\mathrm{rm-1}} = (1-\alpha)(\theta_{\mathrm{od-1}} + \alpha\theta_{\mathrm{od-2}} + \alpha^2\theta_{\mathrm{od-3}} + \ldots), \tag{6.6}$$

with $\theta_{\mathrm{od-i}}$ being the daily mean temperature i days before and α a constant, having a recommended value $\alpha = 0.8$. The CEN standard defines the comfort temperature using a linear regression between reported comfort temperature and the above exponentially weighted running mean outdoor temperature, defined in CIBSE Guide A [123] as:

$$\theta_{\mathrm{comf}} = \begin{cases} 22.6 + 0.09 \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} \leq 10°\mathrm{C}) \\ 18.8 + 0.33 \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} > 10°\mathrm{C}) \end{cases}$$

whose dependence is shown in Figure 6.2(a).

ASHRAE standard. The ASHRAE standard 55-2004 [105] proposes on the other hand a similar model that takes monthly mean outdoor temperature $\theta_{\mathrm{out,mm}}$ as a predictor, based on the observations of De Dear and Brager [124, 125, 126]:

$$\theta_{\mathrm{comf}} = 17.8 + 0.31 \cdot \theta_{\mathrm{out,mm}}, \quad (10°\mathrm{C} \leq \theta_{\mathrm{out,mm}} \leq 33°\mathrm{C}) \tag{6.7}$$

with zones defining predicted 80% and 90% proportions of satisfied (Figure 6.2(b)).

Dynamic detailed human body models

In order to overcome some of the flaws of Fanger's model, another approach taken was to develop detailed dynamic models of human thermoregulation. The first model, proposed by Gagge, Stolwijk and Hardy in 1966 [127, 128], represents the human body as three cylinders (head, trunk and extremities), each of them being divided into at least two concentric layers to account for the anatomical and functional differences of importance in thermoregulation. The model simulates the heat flow between adjacent layers (by conduction), likewise the heat exchanges with the environment (by conduction, convection, radiation and evaporation).

Zhang et al. [129, 130] have since developed a model based on refinements to the approach of Gagge, Stolwijk and Hardy; but now with an unlimited number of body segments, a refined blood flow model, a clothing node (to model heat and moisture capacitance), the inclusion of conductive heat transfer to external contact surfaces, the explicit calculation of radiation heat transfer and a radiation heat flux model.

This approach was also refined by the extremely detailed model proposed by Fiala [131, 132]. In this model the human organism is separated into two interacting systems of thermoregulation: the controlling active system (accounting for regulatory responses such as shivering, vasomotion and sweating) and the controlled passive system (including the heat transfers taking place inside the human organism and at its surface). This latter is based on a representation of the human body as fifteen spherical or cylindrical elements (head, face, neck, shoulders, arms, hands, thorax, abdomen, legs and feet) composed of concentric layers (brain, lung, bone, muscle, viscera, fat and skin) subdivided into tissue nodes.

The models of both Fiala and Zhang have been calibrated to observed thermal sensation and extensively validated for a wide range of thermal conditions (including transient and nonuniform thermal environments) and metabolic activity. These advanced models have been successfully applied to predict comfort based on detailed simulations of the indoor environment, coupled with computational fluid dynamics (CFD) models [133, 134, 135]; a field in which research is now very active.

However, these models need extensive computational resources, such that the wide scale use of these techniques is currently beyond reach. Nevertheless, with its high level of detail this approach offers much future promise for the precise prediction of thermal comfort in the indoor environment (including comprehensive predictions of local and overall sensation, asymmetric conditions and draughts [136, 137]). Nevertheless, as with Fanger's approach the modelling of the effects of occupants' adaptation of their personal environment characteristics is currently ignored.

6.1.2 Visual comfort

A large number of studies report that visual comfort depends mainly on local illuminance, but also on secondary parameters, including illuminance distribution in the indoor space,

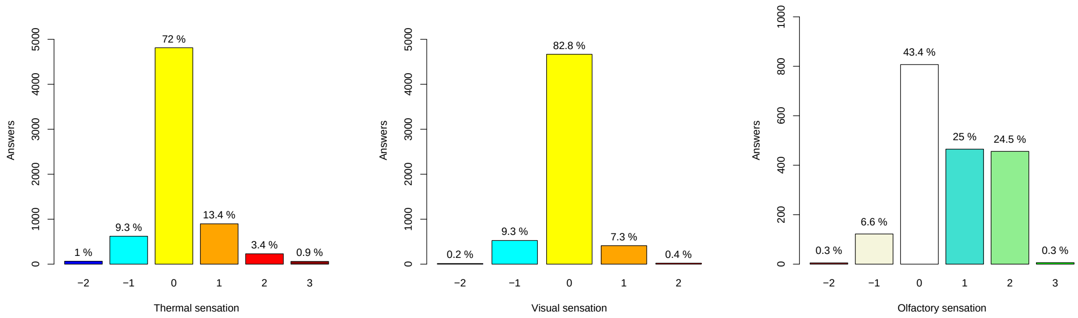


Figure 6.3: Statistical summary of reported thermal (left), visual (centre) and olfactory (right) sensations

colour temperature, local glare and the ability to have a view out. Last but not least, illuminance levels from daylight are not perceived to be equivalent to similar levels provided by electric lighting [138].

A large number of studies have observed that daylight is more comfortable than electric lighting [138], which is not surprising . For instance, Wells [139] observed that 69% of field survey subjects believed that daylight was better for their eyes than electric light and 89% thought that the availability of a view out was important. Research performed by Laurentin et al. [140] observed that subjects perceived an illuminance of 300 lux as pleasant if it arose from daylight and unpleasant in the case of electric lighting.

From field studies significant variations in preferred illuminance levels have also been observed between individuals. In their review of published experimental evidence of preferred illuminance domains, Nabil and Mardaljevic [141] noticed that daylight illuminance below 100 lux is considered insufficient and that 100 to 500 lux is rated as effective. At higher levels of between 500 and 2000 lux its perception varies from desirable to tolerable, and higher values are supposed to create visual discomfort.

Based on a Bayesian analysis of actions on blinds and lighting performed in the LESO building (Section 2.1.1), Lindelöf proposed a non-parametric estimation for the probability of visual discomfort [33, 94]. This latter reaches a global minimum of 30% at about 500 lx and increases moderately until 2500 lx, from which its slope increases sharply.

6.2 Preliminary observations

6.2.1 Statistical summary

We first examine the distribution of observed sensation votes (Figure 6.3) and summarise the overall comfort conditions prevailing in the surveyed building. For this we will use two definitions of thermal and visual comfort: an occupant is considered *strictly comfortable* if the vote falls in the central category, and *loosely comfortable* for a vote among the three central categories (which is the usual definintion of comfort, see Table 6.1) of the bidirectional sensation scales.

Occupants reported to be strictly thermally comfortable on 72.0% of occasions and loosely comfortable for 94.7%. The occurrence of visual comfort is even higher with 82.8%, and up to 99.4% for its loose definition. The distribution of comfort votes is significantly biased towards a warm sensation for thermal comfort, while it is almost symmetric for visual sensation.

The proposed scale for olfactory sensation does not display such a symmetry in its formulation (it is unidirectional) so that the observed distribution is non-symmetric. Only 6.9% qualify it as worse than acceptable, while 49.8% of votes are explicitly positive; although the maximum rating of "Excellent" is very rarely chosen (0.3%).

Individual occupants have reported between 43.7% and 99.0% of their answers to be strictly thermally comfortable (between 84.6% and 100% for loosely comfortable). These proportions range from 34.8% to 96.4% for strict visual comfort (from 96.5% to 100% for loose comfort; it is lower than 100% for only nine occupants).

We have shown in Section 2.3.2 that the questionnaire was administered for a representative subset of the prevailing global indoor thermal and visual conditions, which makes these proportions a valid summary of the overall environmental conditions experienced in the building. The high number of comfortable votes shows that the building generally delivers a good quality indoor environment to its occupants. On the other hand, the small amount of uncomfortable events makes departure from comfort more difficult to detect, particularly when the loose definition of comfort is used. For this reason we will generally study strict comfort in more detail in this chapter, to allow for more precise estimates of (dis)comfort conditions.

6.2.2 Application of the adaptive model

In this section, we check whether our data is in agreement with previously proposed adaptive models for thermal comfort (mentioned in Section 6.1.1). The usual practice for the derivation of their regression coefficients is to perform linear regression between the comfort indoor temperature θ_{comf} and the retained temporal average of θ_{out}, where θ_{comf} is defined as proposed by Griffiths [142] as $\theta_{\text{comf}} = \theta_{\text{in}} - 2S_{\text{th}}$, based on the observation of De Dear and Brager [125] that an increase of temperature of 1°C corresponds to a 0.5 increase in thermal sensation.

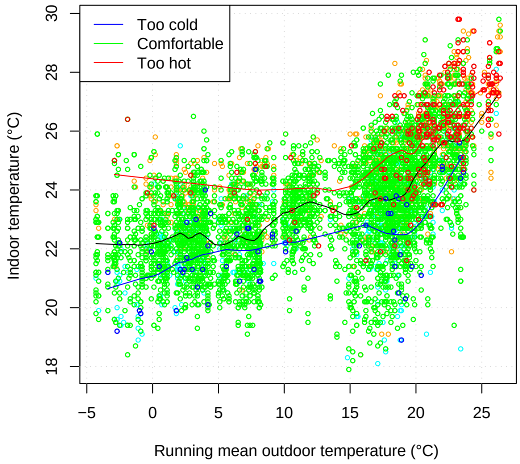
Using an alternative method, we also perform linear regression between θ_{in} and the retained temporal average of θ_{out}, but using the data points corresponding to comfortable votes only. An advantage of this approach is that we remove the assumed relationship between θ_{in} and S_{th}, which allows for the direct use of untransformed data, based on temperatures that are explicitly rated as comfortable. The drawbacks lie in the use of a restricted subset of the database and in the assumption that this subset is representative of the potential comfort conditions[3].

We do not provide a comparison with the predictions from Fanger's model, as some key variables (v_{air}, θ_{mr}) were not surveyed in our experimental campaign and because there is a growing consensus that this model is inapplicable to transient environments (Section 6.1.1).

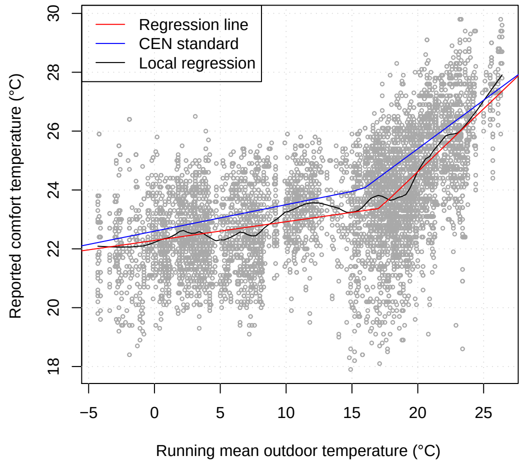
Results based on events rated comfortable

Comparison with CEN standard. Using our data to perform a linear regression between θ_{in} and $\theta_{\text{out,rm}}$ for indoor thermal comfort, as recommended within the CEN stan-

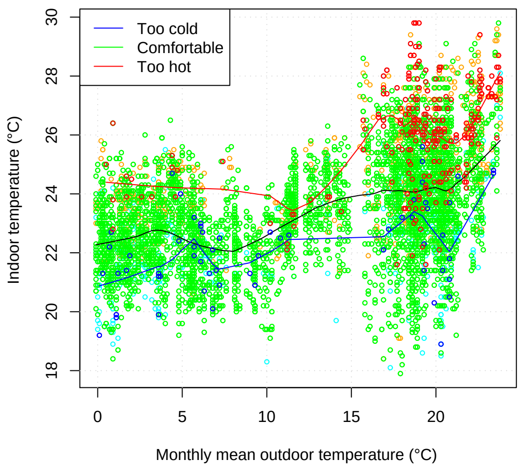
[3]Such a problem may occur when the amount of temperature measurements rated as uncomfortable is insufficient. In this case, there is no guarantee that the subset of data points rated as comfortable is unbiased, so that the regression may be misleadingly based on a non-representative subset of the whole domain of comfortable conditions. However, regression using the method of Griffiths requires similar assumptions.



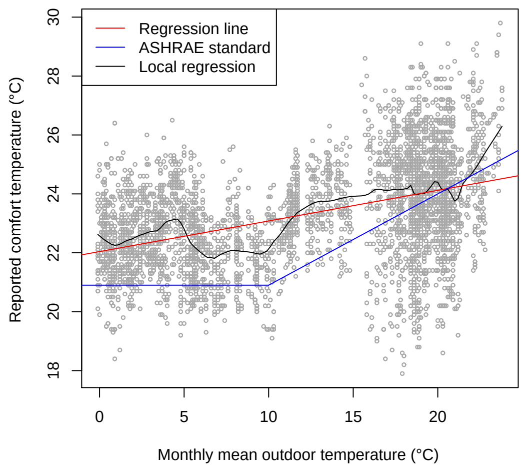
(a) Observed indoor temperature and thermal comfort votes versus running mean outdoor temperature, with local polynomial regression



(b) Observed comfortable indoor temperature versus running mean outdoor temperature with local polynomial and linear regression and comfort line predicted by the CEN standard



(c) Observed indoor temperature and thermal comfort votes versus monthly mean outdoor temperature, with local polynomial regression



(d) Observed comfortable indoor temperature versus monthly mean outdoor temperature with local polynomial and linear regression and comfort line predicted by the ASHRAE standard

Figure 6.4: Comparison with the adaptive models given in the CEN and ASHRAE standards

dard, we obtain:

$$\theta_{\mathrm{in,comf}} = \begin{cases} (22.24 \pm 0.04) + (0.060 \pm 0.005) \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} \leq 10°\mathrm{C}) \\ (16.90 \pm 0.31) + (0.378 \pm 0.016) \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} > 10°\mathrm{C}) \end{cases}$$

Figures 6.4(a)-6.4(b) present both local polynomial and linear regressions as well as the raw data. The local polynomial models the key tendencies in the data and matches closely the partitioned CEN regression equation. Also presented in Figure 6.4(a) are local polynomial regression lines corresponding to strict cold (lower) and hot (upper) discomfort; so that it is visible that a sufficient amount of uncomfortable events bound the comfortably rated data points, thus resolving the potential drawback of only using a subset of the available data.

Comparison with ASHRAE standard. Performing linear regression on $\theta_{\mathrm{out,mm}}$ as proposed in the ASHRAE standard, we obtain using our data:

$$\theta_{\mathrm{in,comf}} = (22.04 \pm 0.04) + (0.104 \pm 0.003) \cdot \theta_{\mathrm{out,mm}}.$$

In this case the derived values for both the intercept and the slope are much lower than those given in the standard. This may be due to climatic and/or building design differences between the two datasets used for the development of these standards. Based on our data, we see in Figures 6.4(c)-6.4(d) that $\theta_{\mathrm{out,mm}}$ is more weakly associated with observed variations in $\theta_{\mathrm{in,comf}}$ than is the case using $\theta_{\mathrm{out,rm}}$.

Analysis based on the method of Griffiths.

As with McCartney and Nicol [66], we deduce the comfort temperature from θ_{in} and S_{th}. Performing linear regression, we obtain for the CEN standard equation (Figure 6.5(a)):

$$\theta_{\mathrm{in,comf}} = \begin{cases} (22.38 \pm 0.04) + (0.046 \pm 0.005) \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} \leq 10°\mathrm{C}) \\ (19.52 \pm 0.28) + (0.230 \pm 0.014) \cdot \theta_{\mathrm{out,rm}} & (\theta_{\mathrm{out,rm}} > 10°\mathrm{C}) \end{cases}$$

Similarly, based on the monthly mean outdoor temperature adopted by the ASHRAE standard: (Figure 6.5(b)):

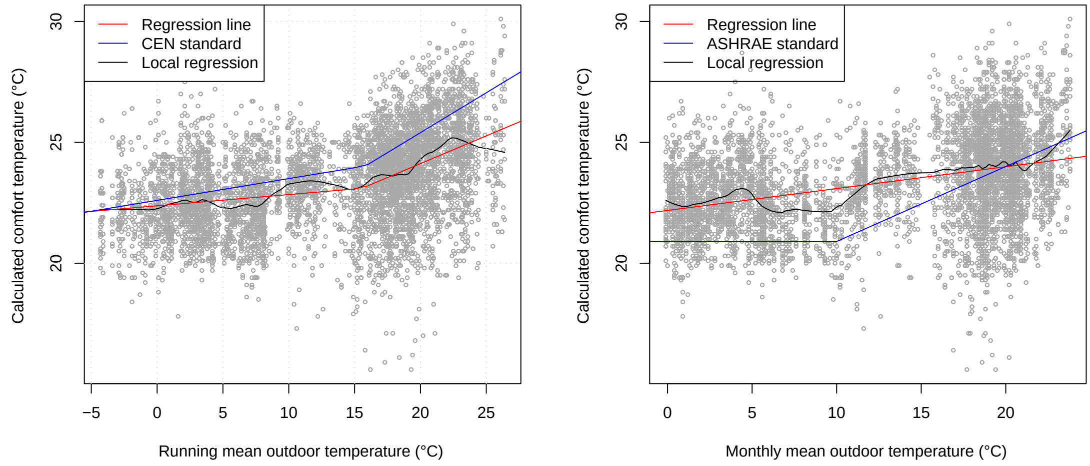
$$\theta_{\mathrm{in,comf}} = (22.18 \pm 0.04) + (0.090 \pm 0.002) \cdot \theta_{\mathrm{out,mm}}.$$

Remarks

From the above observations, it seems clear that $\theta_{\mathrm{out,rm}}$ is a better quality predictor of comfort than $\theta_{\mathrm{out,mm}}$; this is supported by previous research, while the use of monthly means has attracted some criticism[4].

In spite of the good observed agreement with the CEN adaptive comfort model, we do not consider this approach to be fully satisfactory; in particular $\theta_{\mathrm{out,rm}}$ seems to be a simplistic predictor for the variations in comfort temperature, supposed to account for occupants' expectations and adaptive actions. For these latter, the link between θ_{comf} and $\theta_{\mathrm{out,rm}}$ may arise from a spurious correlation, which should be reformulated to take the

[4]For instance, Nicol and Humphreys point out in their recent article [121] that the temperature can be very variable within a month and that the value of monthly mean is subject to misinterpretation.



(a) Deduced comfort temperatures versus running mean outdoor temperature with local polynomial and linear regression and comfort line predicted by the CEN standard

(b) Deduced comfort temperatures versus monthly mean outdoor temperature with local polynomial and linear regression and comfort line predicted by the ASHRAE standard

Figure 6.5: Comparison with the adaptive models given in the CEN and ASHRAE standards, based on the method of Griffiths

actual influencing predictors as explanatory variables, by analysing explicitly the impact of adaptive actions on comfort temperature. In particular, we may expect a different behaviour of θ_{comf} if a smaller or larger range of control actions is available compared with those available within the buildings from which these regression were derived. This issue is discussed in Section 6.4.

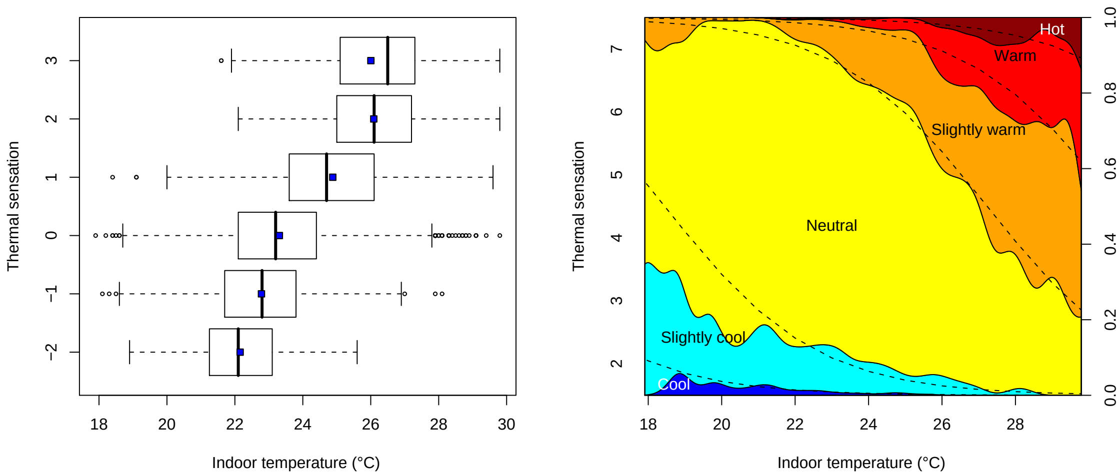
6.3 Predicting thermal comfort

Previous research largely documents that indoor temperature is the most influential variable on thermal comfort (Figure 6.6(a)). In the following we focus on indoor temperature as the driving stimulus for thermal comfort and propose an informative probabilistic approach.

6.3.1 Distribution of thermal sensation

Rather than simply performing a linear regression to predict a mean comfort temperature, it is of greater interest to study the distribution of comfort votes with respect to its driving variable – indoor temperature (Figure 6.6(b)). The large spread observed above suggests the use of a probabilistic approach. We propose then to model the distribution of thermal sensation with respect to indoor temperature as a proportional odds ordinal logistic model for $P(S_{th}|\theta_{\text{in}})$, formulated as

$$p(S_{\text{th}} \geq S_j) = \frac{\exp(a_j + b \cdot \theta_{\text{in}})}{1 + \exp(a_j + b \cdot \theta_{\text{in}})}, \tag{6.8}$$



(a) Box plots of observed indoor temperatures versus reported thermal sensations

(b) Distribution of observed thermal sensation votes conditional on indoor temperature, with fitted ordinal logistic distributions

Figure 6.6: The link between thermal sensation and indoor temperature

where the regression parameters are given in Table 6.3 (left). Equation (6.8) proposes a detailed prediction of thermal sensation, from which the probability of strict comfort $p(S_{th} = 0) = p(S_{th} \geq 0) - p(S_{th} \geq 1)$ is easily deduced; likewise the probability of loose comfort. This also enables a rigorous treatment of the dependent variable – thermal sensation – which is a discrete variable.

Figure 6.6(b) offers a useful graphical summary of the observed and fitted distributions of thermal sensation with respect to indoor temperature, and shows that the fitted ordinal logistic probabilities correctly represent the measured distributions. For any chosen value of θ_{in} on the x-axis, the probability distribution of S_{th} may be read along the associated vertical line.

It can be noticed that the central thermal sensation represents the absolute majority of the outcomes for temperatures between ca. 18°C and 27°C. However, it cannot be excluded that this phenomenon could be caused by the wording that was adopted for the thermal sensation scale, where the proposed central category may cover a wider semantic range than other alternatives, rather than an intrinsic bias of the population of this building (which presents its occupants with many adaptive opportunities) towards thermal neutrality.

6.3.2 Thermal comfort probability

Although such a prediction of thermal sensation distribution informative, it is of more direct interest (in terms of usability) to simply predict a probability of comfort. The drawback of this formulation compared to ordinal probabilities is that we would only examine whether (strict or loose) comfort is achieved, which may seem less comprehensive than a measure like Fanger's PMV. However, the probability of (dis)comfort is directly related to the predicted percentage of dissatisfied (PPD) which is of interest to assess the acceptability of the indoor environment.

Occupant discomfort is defined by the occurrence of the events "being too cold" or

	Thermal sensation		Visual sensation	
	Estimate	**Wald Z**	**Estimate**	**Wald Z**
a_j, $(S_j = -1)$	-6.44 ± 0.38	-16.83	1.00 ± 0.48	2.06
a_j, $(S_j = 0)$	-8.95 ± 0.36	-24.59	-2.64 ± 0.35	-7.48
a_j, $(S_j = +1)$	-13.22 ± 0.39	-33.59	-7.64 ± 0.41	-18.79
a_j, $(S_j = +2)$	-14.95 ± 0.41	-36.89	-10.80 ± 0.52	-20.62
a_j, $(S_j = +3)$	-16.59 ± 0.43	-39.04		
b	0.485 ± 0.016	30.24	0.769 ± 0.057	13.41
Goodness of fit	$AUC = 0.715$, $D_{xy} = 0.43$, $\Gamma = 0.436$ $\tau_a = 0.193$, $R^2_N = 0.173$, $B = 0.009$		$AUC = 0.709$, $D_{xy} = 0.419$, $\Gamma = 0.423$ $\tau_a = 0.133$, $R^2_N = 0.109$, $B = 0.003$	

Table 6.3: Ordinal logistic regression parameters on thermal (left) and (visual) sensation votes

Type	**a**	**Z**	**b**	**Z**	**LR**	**AUC**	$\mathbf{D_{xy}}$	$\mathbf{\Gamma}$	τ_a	$\mathbf{R^2_N}$	**B**
Strict											
p_cold	4.49 ± 0.55	8.1	-0.287 ± 0.024	-11.9	152.40	0.643	0.286	0.290	0.053	0.049	0.090
p_hot	-15.84 ± 0.53	-30.1	0.591 ± 0.021	27.7	969.07	0.779	0.558	0.565	0.161	0.236	0.120
p_dark	2.52 ± 0.42	6.1	-0.750 ± 0.069	-11.0	129.86	0.735	0.471	0.475	0.088	0.111	0.089
p_bright	-3.21 ± 0.13	-24.9	$(6.3 \pm 0.7) \cdot 10^{-4}$	9.0	74.18	0.689	0.378	0.387	0.056	0.073	0.071
Loose											
p_cold	5.70 ± 1.74	3.3	-0.453 ± 0.078	-5.8	36.85	0.727	0.454	0.488	0.009	0.057	0.009
p_hot	-21.21 ± 0.97	-21.8	0.731 ± 0.038	19.4	152.40	0.643	0.286	0.290	0.053	0.049	0.090
p_dark	0.53 ± 1.05	0.5	-1.08 ± 0.21	-5.2	22.46	0.894	0.788	0.859	0.005	0.214	0.004
p_bright	-7.00 ± 0.69	-10.2	$(9.6 \pm 2.8) \cdot 10^{-4}$	3.4	11.13	0.775	0.550	0.669	0.004	0.097	0.004

Table 6.4: Regression parameters and goodness-of-fit estimators for thermal and visual strict and loose discomfort probabilities based on the whole set of surveyed occupants

"being too hot", from the point of view of strict or loose (dis)comfort (Section 6.2.1), to which a probability can be assigned. Using the strict definition of comfort, the probability of being too cold is defined as $p_\mathrm{cold} = \mathrm{Prob}(S_\mathrm{th} < 0)$, while we set $p_\mathrm{hot} = \mathrm{Prob}(S_\mathrm{th} > 0)$ for the probability of being too hot. In both cases, the *discomfort probability* p_discomf is directly obtained from p_cold and p_hot:

$$\mathrm{Prob}(S_\mathrm{th} \neq 0) = \mathrm{Prob}(S_\mathrm{th} < 0) + \mathrm{Prob}(S_\mathrm{th} > 0), \quad p_\mathrm{discomf} = p_\mathrm{cold} + p_\mathrm{hot}, \tag{6.9}$$

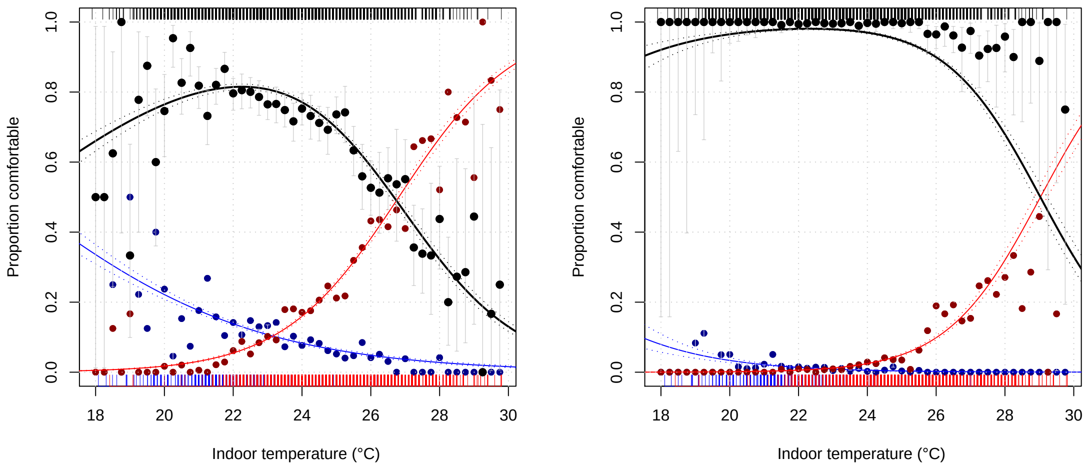
from which the *comfort probability* $p_\mathrm{comf} = 1 - p_\mathrm{discomf} = 1 - p_\mathrm{cold} - p_\mathrm{hot}$ is deduced. However, the associated property $p_\mathrm{cold} + p_\mathrm{hot} + p_\mathrm{comf} = 1$ does not necessarily hold when p_cold and p_hot are separately fitted to the data. In this case, the term $p_\mathrm{cold} \cdot p_\mathrm{hot}$ must be considered, which gives

$$p_\mathrm{discomf} = p_\mathrm{cold} + p_\mathrm{hot} - p_\mathrm{cold} \cdot p_\mathrm{hot} \tag{6.10}$$

$$p_\mathrm{comf} = 1 - p_\mathrm{discomf} = (1 - p_\mathrm{cold}) \cdot (1 - p_\mathrm{hot}). \tag{6.11}$$

These probabilities potentially depend on a large set of environmental variables, the most significant of which should be included in a predictive model. We have chosen to fit logistic distributions to p_cold and p_hot and check for the significance of available variables. We find that indoor temperature is a significant predictor for p_cold and p_hot (Table 6.4) and that the inclusion of further predictors does not significantly improve the quality of fit.

The obtained distributions of p_cold, p_hot and p_comf are displayed in Figure 6.7, for both strict and loose definitions of thermal comfort. From this we observe that the fitted



(a) Strict definition of comfort (b) Loose definition of comfort

Figure 6.7: Fitted thermal comfort probabilities (black line), and discomfort probabilities (blue and red lines) with standard errors (dashed lines) and observed proportions of comfortable votes with their binomial 95% confidence intervals

distributions are in close agreement with observed proportions of comfort and discomfort. In the case of loose comfort, we observe that p_{comf} reaches a maximum of close to 95%, which reproduces the classical observation of Fanger with respect to the PPD, and that it is consistently superior to 90% in the interval 17.5°C to 25.9°C; which does not match with Fanger's near Gaussian distribution of PPD.

The slope associated with p_{cold} is smaller than for p_{hot}, likewise its precision, statistical significance and goodness-of-fit estimators (Table 6.4). This is due to the rather low frequency of cold events, which dampens the accuracy of discomfort predictions in low temperature domain. If more measurements were performed in these conditions we may have expected a sharper increase in discomfort probability (for instance, with the current regression parameters, $p_{cold} = 0.9$ is reached for 8°C, which is not coherent).

The uncertainties on regression parameters are bigger when fitting for loose comfort, which is based on less frequent discomfort events. Furthermore, the result is less informative (the probability distribution is almost flat) as our measurements do not cover effectively enough the temperature ranges where strong discomfort should be prevalent. Finally, slope estimates are larger with loose comfort, as the rare discomfort events are more clearly separated from comfortable data points than with the strict definition.

Compared to the ordinal logistic regression model above, this approach allows for the representation of non-equal slopes in p_{cold} and p_{hot} and thus to include the possibility of an asymmetry in p_{comf}.

6.3.3 Individual behaviours

Être soi-même!... Mais soi-même en vaut-il la peine?
To be oneself!... But is oneself worthy of being?
Paul Valéry (1871-1945), Mauvaises pensées et autres

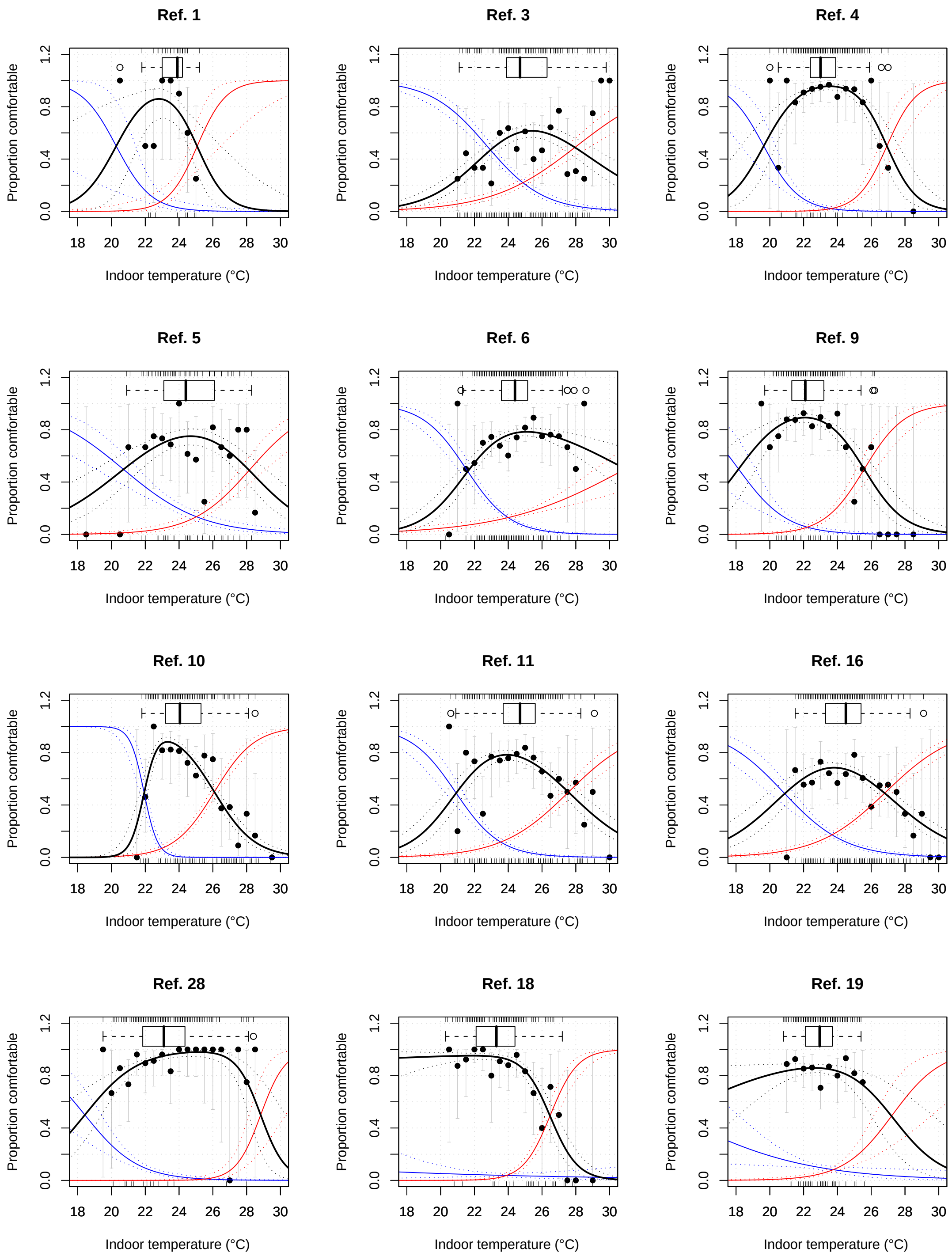


Figure 6.8: Fitted thermal comfort probabilities (black line), and discomfort probabilities (blue and red lines) with fit standard errors (dashed lines), observed proportions of comfortable votes with their binomial 95% confidence intervals and box plots of reported comfort temperatures (part 1)

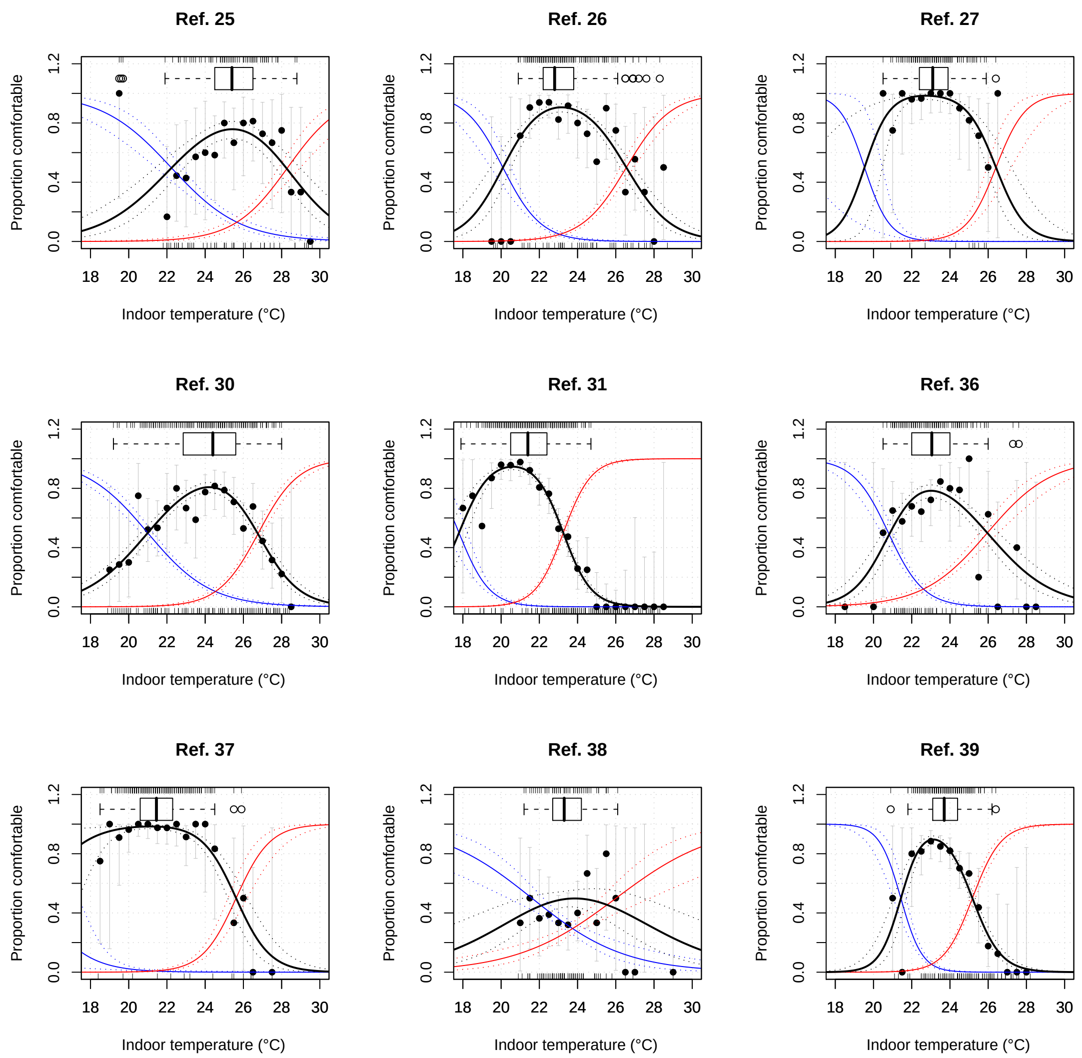


Figure 6.9: Fitted thermal comfort probabilities (black line), and discomfort probabilities (blue and red lines) with fit standard errors (dashed lines), observed proportions of comfortable votes with their binomial 95% confidence intervals and box plots of reported comfort temperatures (part 2)

Ref.	N	N_{cold}	N_{hot}	a_{cold}	b_{cold}	a_{hot}	b_{hot}	$\hat{p}_{comf}$	$\hat{\theta}_{comf}$	$\theta_{50,cold}$	$\theta_{50,hot}$
1	51	4	9	19.62	-0.967	-29.26	1.167	0.860	22.8	20.3	25.1
3	243	65	57	13.58	-0.597	-11.18	0.398	0.616	25.4	22.8	28.1
4	244	13	10	18.68	-0.951	-30.94	1.149	0.957	23.5	19.6	26.9
5	120	22	20	8.83	-0.427	-16.39	0.581	0.751	24.7	20.7	28.2
6	578	72	92	16.57	-0.770	-8.51	0.275	0.784	25.1	21.5	31.0
9	222	15	23	12.64	-0.697	-21.47	0.839	0.892	22.1	18.1	25.6
10	263	15	82	54.30	-2.477	-21.83	0.837	0.883	23.3	21.9	26.1
11	384	38	73	15.70	-0.752	-13.87	0.503	0.784	23.9	20.9	27.6
16	415	58	118	11.27	-0.542	-13.21	0.492	0.685	23.8	20.8	26.9
18	204	8	19	-1.13*	-0.089*	-32.99	1.245	0.952	21.9	NA	26.5
19	248	25	15	3.61	-0.254	-18.82	0.691	0.859	22.6	14.2	27.2
25	138	28	23	12.24	-0.551	-20.19	0.713	0.758	25.4	22.2	28.3
26	313	24	37	20.08	-0.999	-22.74	0.856	0.906	23.2	20.1	26.6
27	177	2	8	29.91	-1.531	-34.22	1.296	0.984	22.7	19.5	26.4
28	214	15	2	11.70	-0.635	-39.72	1.377	0.980	25.2	18.4	28.8
30	410	77	79	12.76	-0.608	-25.66	0.958	0.808	24.2	21.0	26.8
31	661	14	243	22.39	-1.258	-31.79	1.367	0.946	20.6	17.8	23.3
36	276	35	46	22.61	-1.088	-15.46	0.595	0.784	23.0	20.8	26.0
37	283	3	12	11.64	-0.770	-28.88	1.127	0.984	21.2	15.1	25.6
38	224	51	43	8.86	-0.410	-9.98	0.384	0.498	23.9	21.6	26.0
39	429	17	112	38.24	-1.784	-33.70	1.339	0.896	23.1	21.4	25.2
All	**6685**	**686**	**1187**	**4.49**	**-0.287**	**-15.84**	**0.591**	**0.815**	**22.2**	**15.6**	**26.8**

Table 6.5: Occupant specific summary of thermal comfort probability: total number of answers, number of discomfort answers, logistic regression parameters for discomfort probabilities, maximum fitted comfort probability and its corresponding temperature and characteristic temperatures for discomfort. Asterisks are used to identify statistically non-significant parameters.

Method

We observe also the statistical significance of a multi-level factor for individuals in the analysis of deviance of p_{cold} and p_{hot}. We performed thus a similar analysis by splitting the database by occupants, in order to express variations among individuals in terms of parameters of p_{comf}. However, the analysis was not relevant for some of them, mostly because they had not been surveyed for a sufficiently long period, or because the structure of their answers was not coherent. We removed from the analysis the individuals that did not satisfy the following conditions:

- At least two votes of too cold and two votes of too hot (for the convergence of the fit),
- At least 40 answers (for acceptable standard errors on estimates),
- The obtained probability p_{hot} is an increasing function of θ_{in}, and conversely for p_{cold} (for subject-matter considerations).

Six individuals did not meet these requirements and so were removed from the database. The observed individual comfort probability distributions are displayed in Figures 6.8-6.9, and their regression parameters in Table 6.5.

Individual diversity

It can be seen from Figures 6.8-6.9 that our individuals display considerable variability in the domains where they report to be comfortable. Individual comfort probabilities may be determined using the coefficients a_{cold}, b_{cold}, a_{hot} and b_{hot}. However, as the intercepts a do not have a direct physical interpretation, it is more informative to consider *characteristic discomfort temperatures*[5] $\theta_{50,\text{cold}} = -a_{\text{cold}}/b_{\text{cold}}$ and $\theta_{50,\text{hot}} = -a_{\text{hot}}/b_{\text{hot}}$ as a basis for comparison (Table 6.5). These parameters show modest but significant variations amongst individuals: $\theta_{50,\text{cold}}$ has a sample mean of 19.9°C and a standard deviation of 2.2°C (26.8°C and 1.6°C for $\theta_{50,\text{hot}}$).

There is no strong relationship between $\theta_{50,\text{cold}}$ and $\theta_{50,\text{hot}}$; their correlation coefficient is just 0.27. This demonstrates that occupants show differences in the widths of temperature intervals where they are comfortable, which is visible in Figures 6.8-6.9. The characteristic discomfort temperatures θ_{50} and the slopes b show mild correlations of 0.22 and 0.52, which do not indicate a clear dependence between these parameters.

Some occupants display a significant asymmetry in p_{comf} that can be evaluated through the slopes b: comfort decreases more slowly for hot temperatures if $b_{\text{hot}} < -b_{\text{cold}}$ (and conversely if $b_{\text{hot}} > -b_{\text{cold}}$) while $b_{\text{hot}} \cong -b_{\text{cold}}$ denotes symmetry.

Based on these observations, a possible approach to model occupant variability in thermal comfort probability might involve drawing independently $\theta_{50,\text{cold}}$ and $\theta_{50,\text{hot}}$ from appropriate distributions[6]; likewise for b_{cold} and b_{hot}.

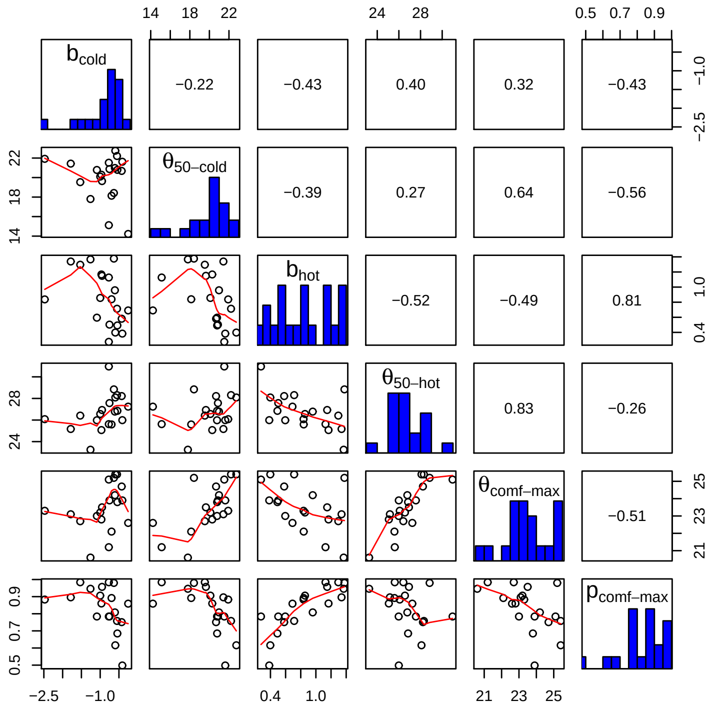


Figure 6.10: Bivariate plots between parameters from individual regression on thermal comfort probability, with local polynomial regression, correlations and histograms

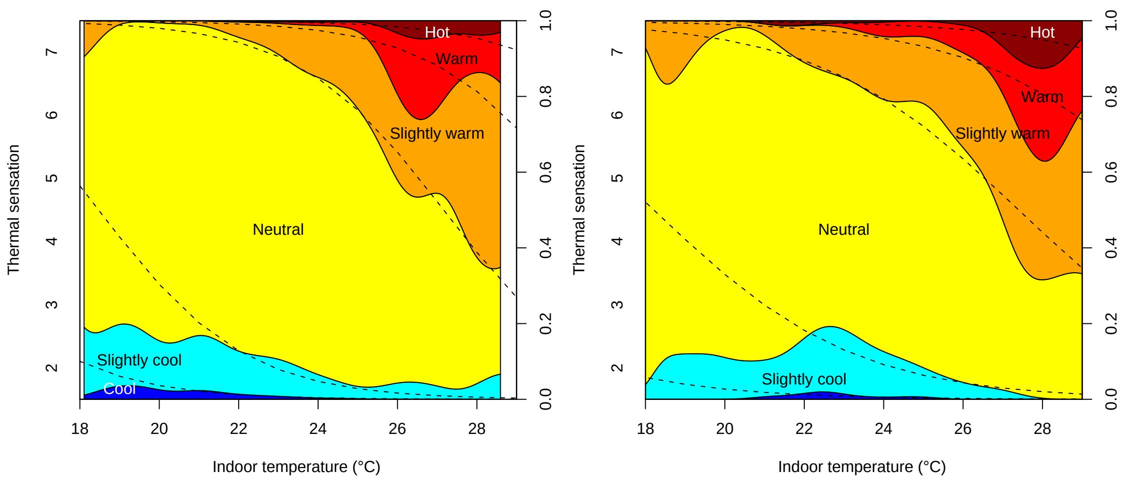


Figure 6.11: Thermal sensation probability distribution with respect to indoor temperature for windows closed (left) and windows open (right)

6.4 Comfort temperature and adaptive actions

We adopt here another approach where we directly examine the distribution of reported comfort temperatures and their dependence on other variables. Having at our disposal simultaneous data for occupants' actions on controls and instantaneous thermal comfort votes, it is of interest to determine whether the use of the studied controls plays a role in the reported comfort temperatures of occupants: whether there is feedback from the adaptive action on perceived thermal sensation and thereby on comfort.

Figure 6.11 shows the distribution of thermal sensation with respect to indoor temperature and the status of windows. The slope associated with θ_{in} lowers from 0.543 ± 0.027 for windows closed to 0.396 ± 0.027 for windows open. This means that when windows are open, our thermal sensation is less sensitive to a temperature increase. For example, the inferred values for $P(S_{th} = 0)$ at $\theta_{\text{in}} = 29°\text{C}$ is 0.266 when windows are closed and 0.333 when they are open.

However an analysis of the effect of adaptive actions on thermal sensation distribution requires a large dataset to obtain reliable estimates, which is achieved here only in the case of actions on windows. Furthermore, it is of more interest and of direct applicability in our work to consider rather the effect of actions on comfort temperature than on thermal sensation.

Figure 6.12 shows the distribution of indoor temperature with respect to reported thermal sensation, distinguishing whether a given adaptive action is exercised (either alone or possibly in conjugation with others). In many cases we observe significant offsets in the mean temperature at which a given thermal sensation vote is reported. It is of particular interest that there are positive offsets where windows are opened and where fans are switched on, for the central sensation vote[7], which are significant according to the two sample t-test ($p < 0.001$). In other words for the same vote occupants tolerate higher temperatures when they have exercised these forms of adaptive control action!

6.4.1 Effects of adaptive actions on thermal sensation: empirical adaptive increments

To ascertain the value added from adaptive actions we determine the difference in mean temperature for neutral thermal sensation votes with and without having exercised a given adaptive action. This is equivalent to the notion of adaptive increments proposed by Baker and Standeven [143], and later reformulated by Oseland et al. [144], albeit based on assumed boundary condition changes to the steady state Fanger model.

The corresponding increments may be read (approximately) from Figure 6.12; they vary from 1.2°C for windows to 1.8°C for fans. However, these latter are generated for cases when the control action in question has or has not been exercised, but others may have been. To precisely estimate the influence of a given control action, the data should be filtered to cases of no significant control action and exclusively the control action in question.

[5] The temperature at which $p_{\text{cold}} = 0.5$ (or $p_{\text{hot}} = 0.5$), see Section 3.2.1, and thus the temperature above which the state of discomfort becomes more likely than comfort.

[6] Normality seems to be a reasonable assumption for characteristic discomfort temperatures, as assymmetry or heavy tails are very unlikely in this context. However, measurements on a larger set of occupants would be needed to draw a clear conclusion using statistical tests for normality.

[7] As in Section 6.2.2, this approach needs the assumption that the temperatures rated as comfortable form a representative set of the whole range of comfort temperatures.

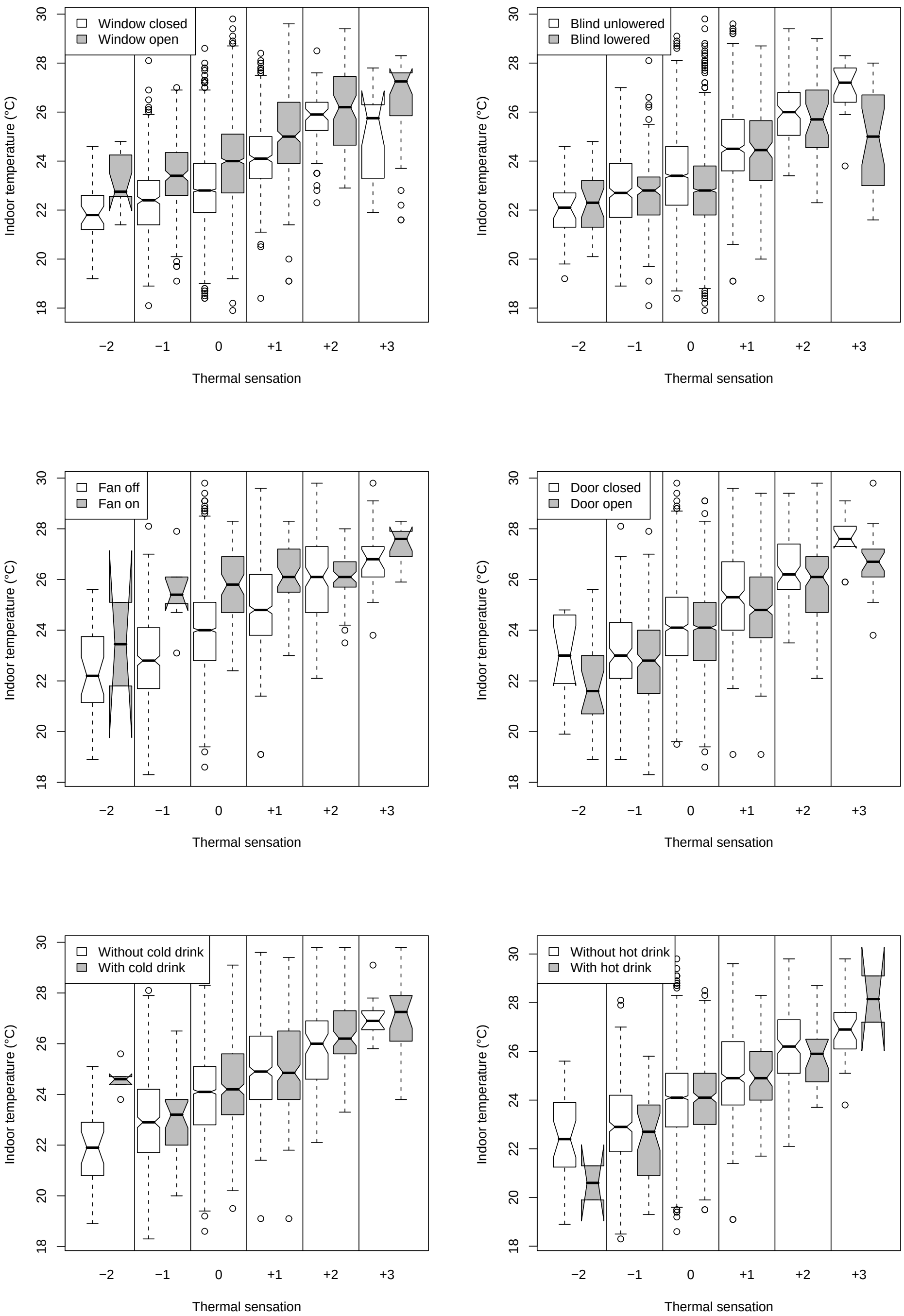


Figure 6.12: Controls' use influence on indoor temperature distribution for given thermal comfort votes

Modelling framework

A rigorous method to evaluate these adaptive increments is to fit a linear model for the prediction of the comfort temperature θ_{comf}, formulated as

$$\theta_{\text{comf},ij} = \mu + \beta_i + \varepsilon_{ij}, \tag{6.12}$$

where $i = 1, \ldots, I$ and $j = 1, \ldots, J_i$; I is the number of considered adaptive actions and J_i the number of observations per level. Setting the case "without" action as the reference group, μ can be interpreted as the mean comfort temperature when no action is taken and the values obtained for β_i are the effects on the comfort temperature produced by a given adaptive action, with ε_{ij} being the fit residuals, assumed to be normally distributed with a mean of zero. With this convention, the values obtained for β_i can then be directly interpreted as the empirical adaptive increments in comfort temperature produced by the associated adaptive action. Furthermore, this convention on the reference group allows for a direct application of related statistical tests, in order to select adaptive actions that have a significant impact on θ_{comf}.

Variable selection

Some adaptive actions do not have a binary nature like the opening and closing of windows, switching on and off fans or consumption or not of drinks, but can rather be exercised incrementally or on a continuous basis. This is the case of clothing, activity and the use of blinds, for which we display the distributions of reported comfort temperatures for their respective levels (Figure 6.13). These actions can be integrated in Equation (6.12) either as continuous variables or as factors with multiple levels.

As expected from its constrained nature, the addition of activity and the status of blinds in this adaptive comfort model is not statistically significant (whether included as multiple level factors or as continuous variables), likewise the binary actions on doors and the intake of cold and hot drinks. However, the current clothing level and actions on windows and fans are significant and are thus retained in the final model.

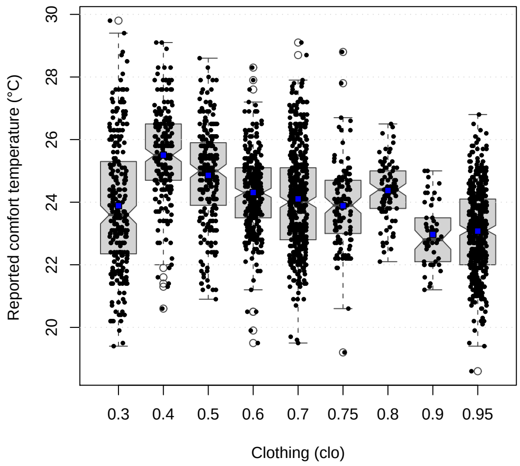
We include clothing level as a transformed variable $\Delta I = I_{\text{cl,max}} - I_{\text{cl}}$ where $I_{\text{cl,max}} =$ 0.95 clo at its maximum. With this convention ΔI measures the degree of clothing removal which is a better definition of clothing adaptation as an adaptive action.

We tested for the significance of interactions between adaptive actions as additional terms in the model. However, no two-term interactions were found to be significant, which may simply be due to inadequate sample sizes for combinations of actions, as estimated standard errors are large. Therefore, we cannot provide clear evidence that these interactions do or do not have an appreciable impact on results[8].

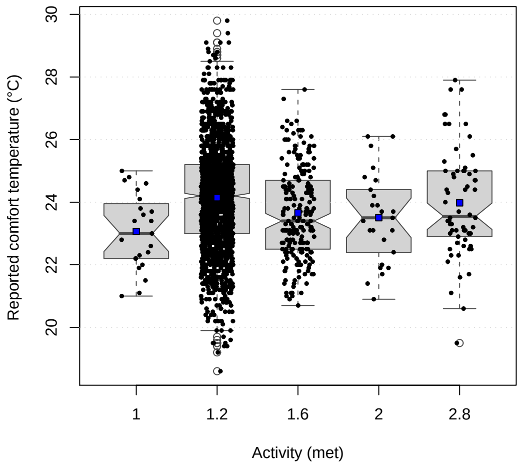
Final model for adaptive actions

Our final model retains thus W, F and $\Delta I = I_{\text{cl,max}} - I_{\text{cl}}$ as parameters, with $R^2 = 0.085$ and $F = 54.21$ ($p < 0.001$). The regression parameters and the analysis of variance are provided in Table 6.6. Our results are graphically summarised in Figure 6.14 with box

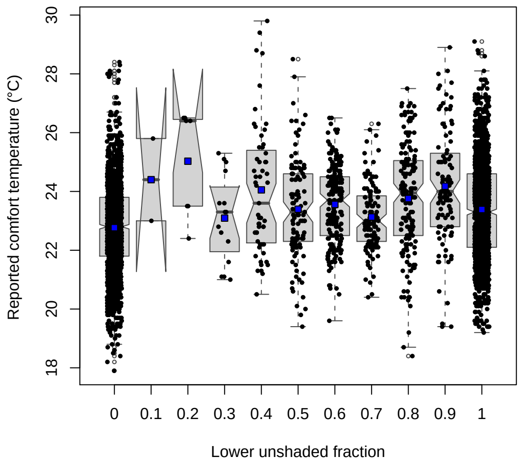
[8]A full list of adaptive increments and their interactions with reasonable uncertainty would require a much larger dataset, or imply a strictly designed experiment with constrained control actions, which is not the meaning of an experiment on free adaptation to the indoor environment. For instance, as fans were never used without other concomittant actions, this increases the uncertainty of its associated estimate compared to a designed experiment.



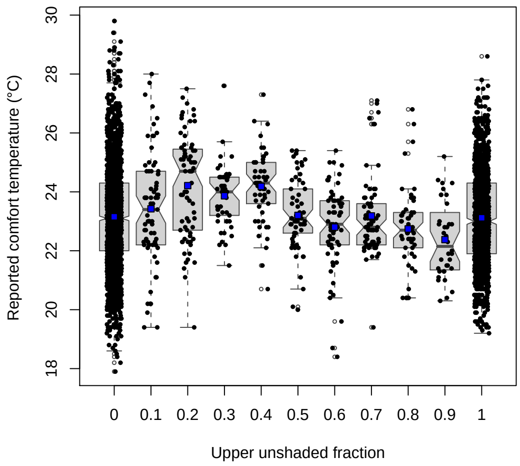
(a) Adaptive increments from clothing



(b) Adaptive increments from activity



(c) Adaptive increments from binned levels of lower unshaded fraction



(d) Adaptive increments from binned levels of upper unshaded fraction

Figure 6.13: Adaptive increments for factors with multiple levels

	Regression parameters			ANOVA				
Variable	Estimate	t	p-value	DF	SS	MS	F	p-value
a	23.29 ± 0.06	370.13	< 0.001					
W	0.50 ± 0.08	6.17	< 0.001	1	245.2	245.2	104.7	< 0.001
F	0.70 ± 0.33	2.13	0.034	1	15.0	15.0	6.4	0.012
$\mathbf{\Delta I}$	1.44 ± 0.20	7.18	< 0.001	1	120.8	120.8	51.6	< 0.001
Residuals				1724	4038.8	2.3		

Table 6.6: Parameters estimates and analysis of variance from the linear model for adaptive actions

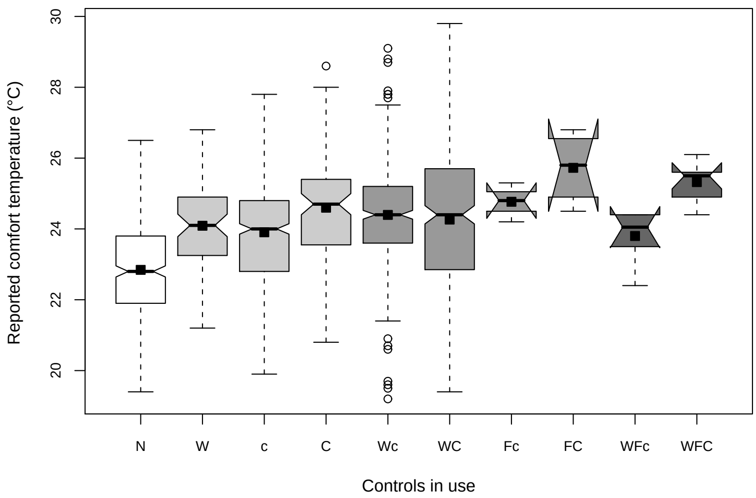


Figure 6.14: Joint influence of significant actions on comfort temperature, with notches along the medians denoting their statistical uncertainty, and solid squares for means (W: windows, F: fans, C: low clothing level (≤ 0.5 clo), c: intermediate clothing level (between 0.6 and 0.8 clo))

plots of observed comfort temperatures for all significant combinations of controls (where the absence of evidence for constructive interactions between controls is visible), and in Table 6.6 with the numerical estimates of the associated increments. It can be observed in Figure 6.14 that our data do not show any signs of skewness or unequal variance between groups. Moreover normal quantile and residuals plots, which are not displayed here, do not show any evidence to limit the scope of application of this linear model.

Finally it is of interest to compare these empirical adaptive increments with those of Oseland et al. [144], who predicted increments of 1.1 °C for windows and 2.2-2.8 °C for fans. These proposed increments are all larger than we have observed.

6.4.2 Individual variations and expanded model

> Chaque homme porte la forme entière de l'humaine condition.
> *Every man bears the whole stamp of the human condition.*
> Michel de Montaigne (1533-1592), Essays (III.2)

Individual comfort temperatures. Variations in the comfort temperature among building occupants are significant (Figure 6.15), in particular with respect to those that were not surveyed in a particular season. A linear model $\theta_{\text{comf},ij} = \mu + \alpha_i + \varepsilon_{ij}$ taking the individual occupant as a unique factor α_i explains a significant part of the observed variance in comfort temperature ($R^2 = 0.372$).

Taking the overall mean comfort temperature as a reference, the regression parameters are simply the offsets of the mean comfort temperatures for each occupant (displayed as

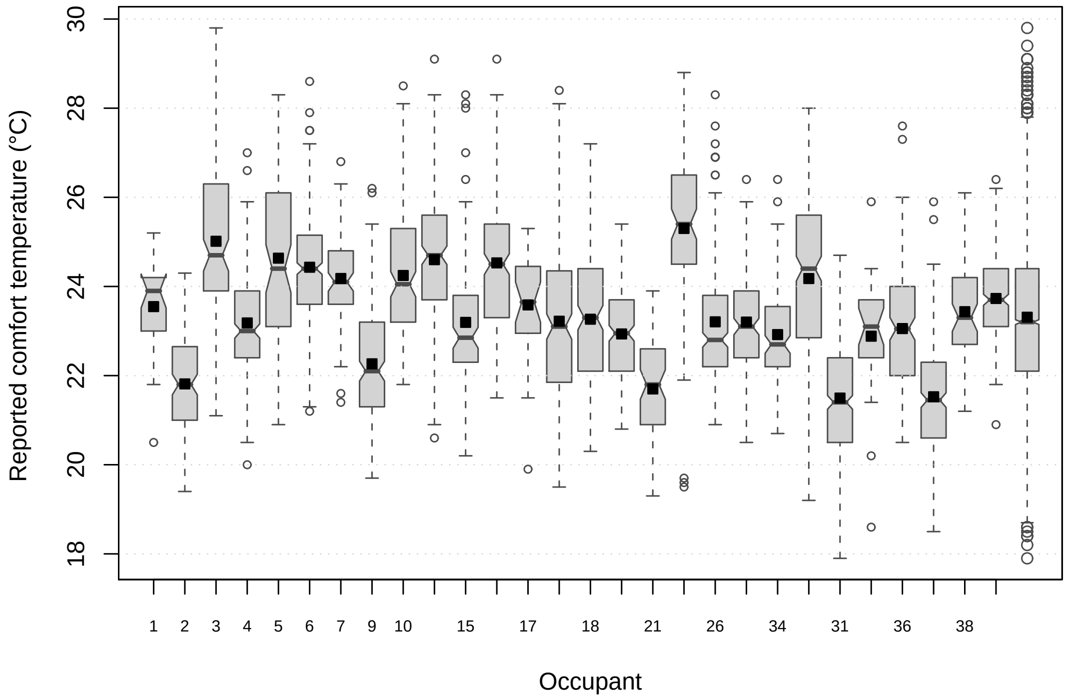


Figure 6.15: Comfort temperature distribution by occupants, with mean values shown as solid squares; the rightmost box plot represents the distribution of aggregated data from the whole set of occupants

black squares in Figure 6.15) with intercept $a = 23.40 \pm 0.03$. These parameters range from -1.91°C (Ref. 23) to +1.90°C (Ref. 18), which is a variation of a similar amplitude as observed for $\hat{\theta}_{\text{comf}}$, $\theta_{50,\text{cold}}$ and $\theta_{50,\text{hot}}$. With a model including together factors for individuals and adaptive actions we observe an improved quality of fit to $R^2 = 0.488$, but with significantly increased standard errors.

Extended model for comfort temperature. Finally, we fit a model adding $\theta_{\text{out,rm}}$:

$$\theta_{\text{comf},ijk} = \mu + \alpha_i + \beta_j + \gamma\theta_{\text{out,rm}} + \varepsilon_{ijk}, \tag{6.13}$$

where $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k = 1, \ldots, K_{ij}$; N is the number of occupants, J the number of considered adaptive actions and K_{ij} the number of observations with respect to occupant i and adaptive action j. With this model the coefficient of determination reaches $R^2 = 0.574$. Thus the proportion of explained variance in the comfort temperature exceeds 50% with the inclusion of adaptive actions, individual differences and $\theta_{\text{out,rm}}$ (to be compared with $R^2 = 0.226$ if $\theta_{\text{out,rm}}$ is the sole predictor).

Variable	Coef.	DF	SS	MS	F	p-value
Windows	β_1	1	245.20	245.20	224.7	< 0.001
Fans	β_2	1	14.96	14.96	13.7	< 0.001
Clothing	β_3	1	120.84	120.84	110.8	< 0.001
Individuals	α	16	1799.15	112.45	103.1	< 0.001
Season	γ	1	377.20	377.20	345.7	< 0.001
Residuals	ε	1707	1862.41	1.09		

Table 6.7: Analysis of variance from the linear model for adaptive actions, individuals and exponentially weighted running mean outdoor temperature

The analysis of variance associated with this extended model is shown in Table 6.7. It is striking that with this model the regression parameter associated with $\theta_{\text{out,rm}}$ is strongly reduced to $\gamma = 0.099 \pm 0.005$, and that its contribution in the analysis of variance is similar to that of adaptive actions and much smaller than individual specificities. As this model includes many predictors in comparison with available data, the associated standard errors are inflated and the parameters' stability is reduced. However, this does not undermine its usability and obtaining more precise parameters is simply a matter of database size.

Results from the method of Griffiths. A similar analysis may be carried out using the assumption of Griffiths (see Section 6.2.2) on the relationship between temperature and thermal sensation, by defining the comfort temperature as $\theta_{\text{comf}} = \theta_{\text{in}} - 2S_{\text{th}}$. The model including adaptive actions only gives $\mu = 23.30 \pm 0.05$, $\beta_W = 0.53 \pm 0.07$, $\beta_F = -0.88 \pm 0.18$ and $\beta_{\Delta I} = 0.81 \pm 0.17$ with $R^2 = 0.047$. This approach produces similar results; noteworthy changes include the change of sign of β_F, which was itself weakly significant using the alternative approach, with a smaller fitted value of $\beta_{\Delta I}$. The fit of the extended model for comfort temperature leads to comparable conclusions; with this convention we obtain $R^2 = 0.418$ and $\gamma = 0.084 \pm 0.006$. Analysis of variance also supports these conclusions.

Remarks. As a conclusion, we underline that this model offers significant improvements compared to previously published adaptive comfort models:

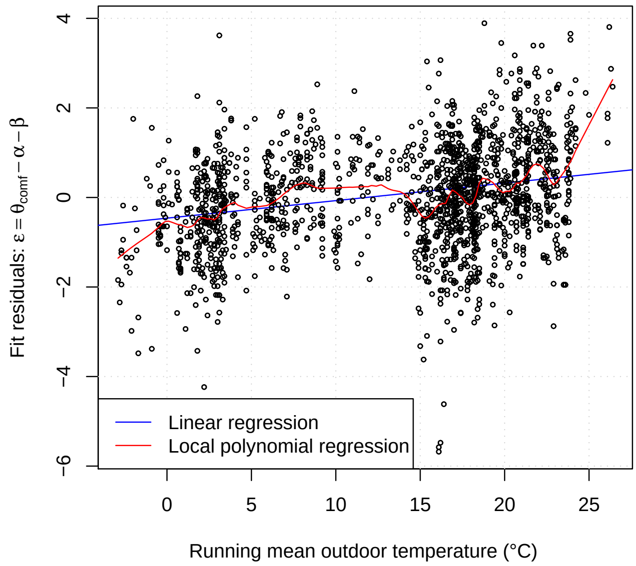


Figure 6.16: Residual influence of $\theta_{\text{out,rm}}$ on θ_{comf}, after fitting for adaptive actions and individual specificities

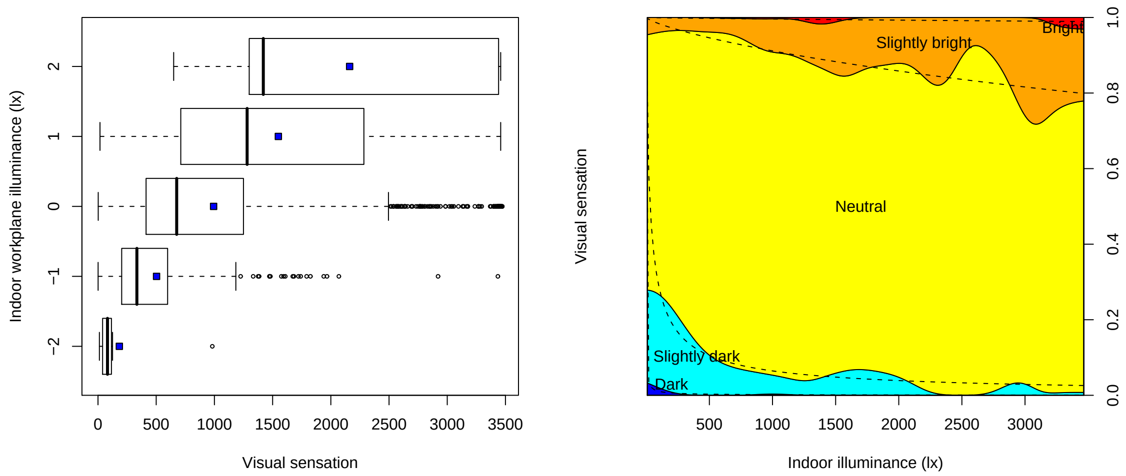
- It separates the contribution of each key predictor for comfort temperature.

- It puts a seasonal variable such as $\theta_{\text{out,rm}}$ in its appropriate place, which accounts to a certain extent for seasonal acclimatisation caused by changes in occupants' expectations of the indoor environment, or other unexplained factors. Figure 6.16 shows that the slope associated with $\theta_{\text{out,rm}}$ on residuals after fitting for actions and individuals is strongly reduced.

- It allows for a calculation of comfort temperature based on available adaptive actions in a particular situation (eg. if clothing adaptation is possible but actions on windows are not).

However, more data (from other buildings and climates) are needed to strengthen the generality of these conclusions and provide reliable parameters for general application of this model. Nevertheless, in spite of the need for further data this approach opens interesting perspectives for the future evolution of adaptive comfort standards as explicit consideration of the weight of adaptive actions is given, which therefore makes this model adjustable to situations where only a subset of possible actions is available.

It is particularly interesting to underline that with this approach not only the slope associated with $\theta_{\text{out,rm}}$ is strongly reduced, but also that on the lower domain of $\theta_{\text{out,rm}}$ the gradient is almost inexistent, while it is strongly dampened for high values of $\theta_{\text{out,rm}}$ (Figure 6.16).



(a) Box plots of observed indoor workplane illuminances versus reported visual sensations

(b) Distribution of observed visual sensation votes conditional on indoor workplane illuminance, with fitted ordinal logistic distributions

Figure 6.17: The link between visual sensation and indoor illuminance

6.5 Predicting visual comfort

In the previous section we studied the linear link between thermal sensation and indoor temperature. Here we apply a similar rationale to attempt to infer a relationship between indoor illuminance and visual sensation. However in this case, we observe that measured illuminance increases exponentially rather than linearly with respect to reported visual sensations (Figure 6.17(a)). As for thermal sensation, we propose to treat their relationship through a probabilistic rationale.

6.5.1 Distribution of visual sensation

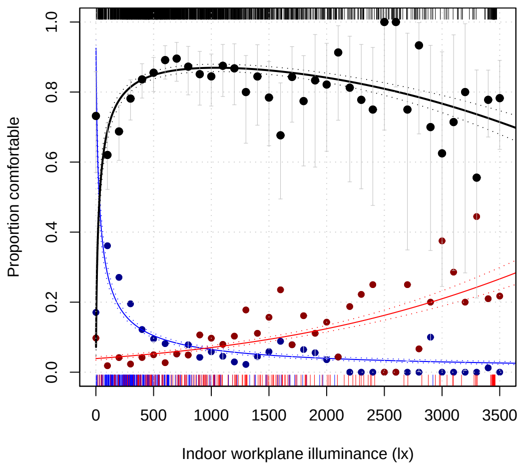
In a similar fashion to our modelling of thermal sensation, we also model the distribution of visual sensation (Figure 6.17(b)) with respect to indoor illuminance, as an ordinal logistic model for $P(S_{\text{vis}}|E_{\text{in}})$,

$$p(S_{\text{vis}} \geq S_j) = \frac{\exp(a_j + b \cdot \log(E_{\text{in}}))}{1 + \exp(a_j + b \cdot \log(E_{\text{in}}))}, \tag{6.14}$$

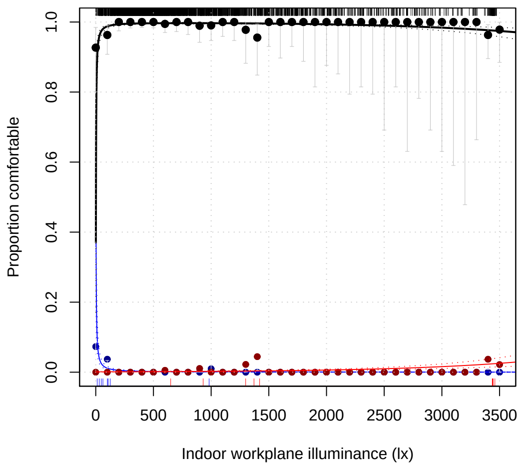
with regression parameters given in Table 6.3 (right). The logarithmic transformation of E_{in} forces the assignment of negative visual sensations for low illuminances, a reasonable outcome even though such dark conditions were not surveyed by our questionnaire. It is noteworthy that neutral visual sensation dominates for the entire surveyed range of illuminances.

6.5.2 Visual comfort probability

We infer a probability distribution for visual comfort using the same principles that we applied to thermal comfort. In this case, the probabilities of being too dark p_{dark} and



(a) Strict comfort



(b) Loose comfort

Figure 6.18: Fitted visual comfort probabilities (black line), and discomfort probabilities (blue and red lines) with standard errors (dashed lines), observed proportions of comfort votes with their binomial 95% confidence intervals

being too bright p_{bright} (using either a strict or loose definition) are estimated, from which *discomfort* and *comfort probabilities* p_{comf} are obtained:

$$p_{\text{discomf}} = p_{\text{dark}} + p_{\text{bright}} - p_{\text{dark}} \cdot p_{\text{bright}} \tag{6.15}$$

$$p_{\text{comf}} = 1 - p_{\text{discomf}} = (1 - p_{\text{dark}}) \cdot (1 - p_{\text{bright}}). \tag{6.16}$$

We observe as expected that E_{in} is a significant predictor for both logistic models and that no other variable brings a significant contribution to p_{dark} and p_{bright}. The obtained probabilities are displayed in Figure 6.18. The transformed variable $\log(E_{\text{in}})$ was used as a predictor for p_{dark}, as the resulting quality of fit is higher and the observed proportions of strict discomfort are better described using this transformation.

The extremely low frequency of visual sensation votes outside the three central categories implies that the obtained probability remains almost flat for the whole range of the surveyed values of E_{in}. The strict probability of comfort is thus of much greater interest. We observe from this definition that p_{comf} reaches a maximum of 0.87 for $E_{\text{in}} = 1090$ (lx).

It is of interest to compare this probability of visual comfort to that one inferred by Lindelöf [33, 94] through Bayesian analysis of actions on blinds and lighting. This analysis led to the different conclusions that the comfort probability[9] reaches a maximum at about 500 lux and that this probability never exceeds 0.7. This emphasises the fundamental difference between discomfort probability and action probability, an issue discussed in detail in Section 6.6.

Finally, we did not observe any visual equivalent to the thermal adaptive increments introduced in Section 6.4, based on the analysis of comfort illuminances with respect to blinds and lighting status, outdoor illuminance and outdoor global and diffuse irradiance.

[9]The author defines it as *"the probability that the user would judge the visual environment as uncomfortable if prompted to do so"*.

Ref.	N	N_{dark}	N_{bright}	a_{dark}	b_{dark}	a_{bright}	b_{bright}	$\hat{p}_{comf}$	$\hat{E}_{comf}$	$E_{50,dark}$	$E_{50,bright}$
3	243	16	3	5.907	-1.253	-10.91	$6.653 \cdot 10^{-3}$	0.927	976	111.7	1640
4	244	6	5	16.036	-3.252	-261.068	$75.11 \cdot 10^{-3}$	0.999	3282	138.6	3476
6	578	51	2	14.305	-2.662	-7.556	$1.413 \cdot 10^{-3}$	0.990	1634	215.8	5348
10	263	14	27	-0.013	-0.388	-2.613	$1.623 \cdot 10^{-3}$	0.808	250	1.0	1610
11	384	29	22	1.556	-0.633	-3.327	$0.302 \cdot 10^{-3}$	0.905	1608	11.7	11008
16	415	81	43	9.581	-1.816	-4.414	$1.424 \cdot 10^{-3}$	0.906	1068	195.8	3100
26	313	30	5	3.698	-0.842	-4.270	$0.602 \cdot 10^{-3}$	0.895	1990	80.9	7088
30	410	45	54	7.542	-1.437	-3.756	$2.123 \cdot 10^{-3}$	0.787	764	190.4	1769
All	**6851**	**538**	**433**	**2.515**	**-0.750**	**-3.207**	$\mathbf{0.627 \cdot 10^{-3}}$	**0.869**	**1044**	**28.5**	**5116**

Table 6.8: Occupant specific summary of visual comfort probability: total number of answers, number of discomfort answers, logistic regression parameters for discomfort probabilities, maximum fitted comfort probability, comfort illuminance and characteristic illuminances for discomfort

However, when less adequate lighting is used, the effect of increased discomfort from artificial lighting noticed by Laurentin et al. [140] could possibly be modelled by "adaptive decrements" generated by the absence of daylight.

6.5.3 Individual behaviours

We applied similar criteria to select individual occupants in order to study their specific probabilities of discomfort, based on the statistical significance of their specificities. Unfortunately in this case, a much smaller number of individual datasets (just eight) meet our selection requirements from the point of view achieving an acceptable quality of fit. The comfort probabilities for our nine retained occupants are presented in Figure 6.19, for whom the regression parameters are presented in Table 6.8.

Occupants clearly report with very similar probabilities that their environment is uncomfortably dark; which contrasts with significant variability with respect to the perception of being uncomfortably bright, some of them being comfortable at high illuminances (Ref. 4, 6, 11, 26 and 38, like many other discarded occupants) while other are more sensitive to glare (Ref. 3, 10, 16 and 30), which causes an increase in p_{bright} at intermediate values of E_{in}.

6.6 Linking actions and comfort

We have studied in detail the probability of action on windows (Chapter 4) and shading devices (Chapter 5), the probability of thermal and visual comfort and the factors influencing comfort temperature. In this section, we show how these concepts are inter-related in a general formulation of human adaptive actions as a response to environmental discomfort. In Section 6.6.1 we show how discomfort leads to action, while we discuss in Section 6.6.2 the feedback to perceived comfort in response to occupants' actions.

6.6.1 A general formulation of human adaptive actions

The notion of action inertia. Figure 6.20(a) shows the probability of thermal discomfort deduced from p_{comf} together with the probability of actions on windows at arrival and

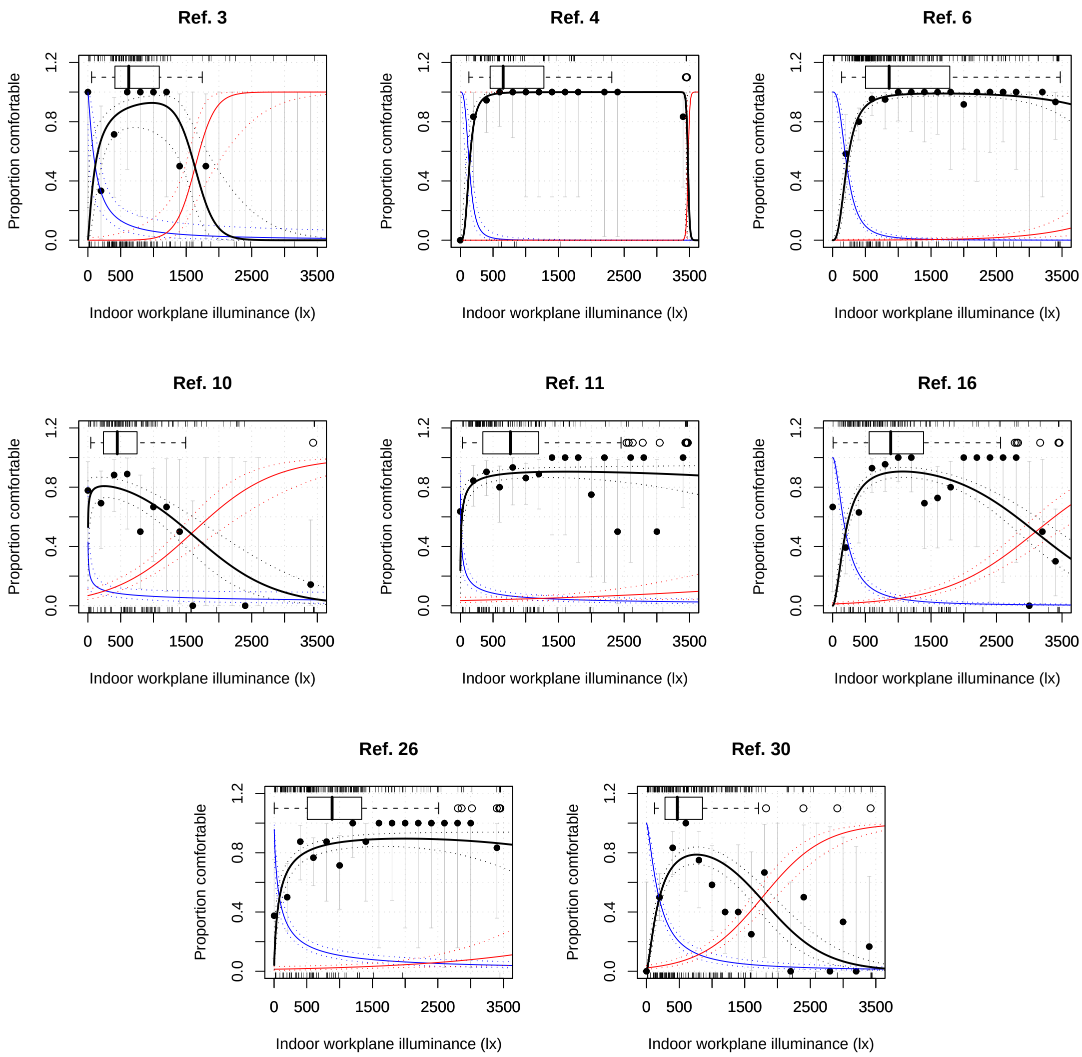
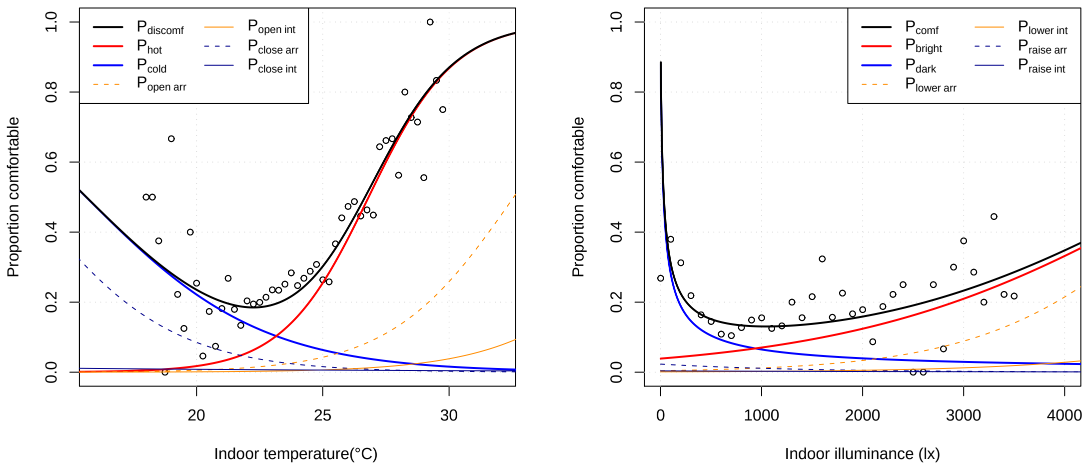


Figure 6.19: Fitted visual comfort probabilities (black lines) and discomfort probabilities (blue and red lines) with fit standard errors (dashed lines), observed proportions of comfortable votes with their binomial 95% confidence intervals and box plots of reported comfortable illuminances



(a) Thermal discomfort probability with probabilities of action on windows

(b) Visual discomfort probability with probabilities of action on blinds

Figure 6.20: Comparing discomfort and action probabilities

during presence[10], as derived in Chapter 4. These curves appear to follow a common trend, although the fitted action probabilities are lower than the comfort probabilities.

From the observation that as discomfort probability increases, so does action probability, we may hypothesise that discomfort causes action, and we may regard the offset between the two as an inertia towards action, a concept proposed by Robinson [69]. We may define this action inertia through some formulation of the discrepancy between action probability and discomfort probability, which can be expressed for example as:

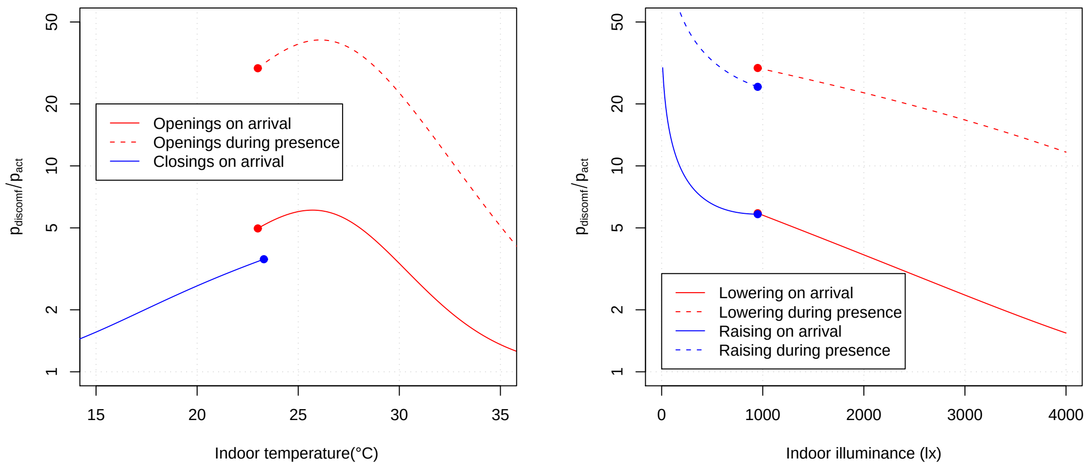
$$I(\theta_{\text{in}}) = \frac{p_{\text{discomf}}(\theta_{\text{in}})}{p_{\text{act}}(\theta_{\text{in}})}. \tag{6.17}$$

This temperature-dependent estimate of inertia is rather complex, but it is both informative and offers a direct link between discomfort and action, as we directly obtain $p_{\text{act}}(\theta_{\text{in}}) = (1/I(\theta_{\text{in}})) \cdot p_{\text{discomf}}(\theta_{\text{in}})$.

Action inertia with respect to thermal comfort. In the case of actions on windows, this inertia is different between arrival (lower) and during presence, which corresponds to the observed higher reactivity in this special case. Figure 6.21(a) shows that as $p_{\text{discomf}}(\theta_{\text{in}})$ increases the action inertia decreases which is an expected result, while it reaches a maximum around $\theta_{\text{in}} = 26°\text{C}$.

The intermediate probability of closing actions is based on outdoor temperature, which undermines the meaning of action inertia, as it implies two different variables; so we consider this no further. We do not study the action inertia for the case of actions on departure, as they are based on predictive rationale rather than on a reaction to recently experienced thermal discomfort.

[10] For the purposes of this analysis, we consider actions probabilities based on the single most influential variable: θ_{in} for $P_{01,\text{arr}}$, $P_{10,\text{arr}}$ and $P_{01,\text{int}}$, and θ_{out} for $P_{10,\text{int}}$.



(a) Temperature-dependent inertia for actions on windows (b) Illuminance-dependent inertia for actions on blinds

Figure 6.21: Behaviour of the stimulus-dependent action inertia, shown on a logarithmic scale

We notice in Figure 6.21(a) that the behaviour of action inertia is very similar between actions on arrival and during presence, with higher values in this latter case; consistent with observed behaviours. Action inertia reaches a maximum at a temperature close to 26°C, above which it decreases, corresponding to a higher reactivity to discomfort as temperature rises.

Action inertia with respect to visual comfort. A similar reasoning can be applied to study the link between visual comfort and actions on blinds. Figure 6.20(b) shows the probability of visual discomfort with the probabilities of actions on blinds (derived in Chapter 5, using E_{in} as the sole predictor).

It is noticeable that occupants display significant inertia in relatively dark conditions, when they are able to switch on electric lighting to counteract visual discomfort. Therefore, a comprehensive estimation of inertia towards visual stimuli needs also to account for actions on lights. This behaviour may also be caused by the relatively low rate of change of illuminance, as the sky darkens relatively gradually.

We observe a regular decrease of inertia as the illuminance rises (Figure 6.21(b)). It is striking that there is almost continuity between inertia towards dark and bright stimuli when the dominant contribution changes. If switch-on probabilities for lighting were available, we could expect to obtain a strong decrease in action inertia for low illuminances, as for p_{dark}.

6.6.2 Comfort feedback of adaptive actions

We have observed that adaptive actions lead to a shift in comfort temperature that allows occupants to feel comfortable in hotter or colder conditions than would be the case if no action was exercised. In order to account for this phenomenon, we can define an *adaptation-*

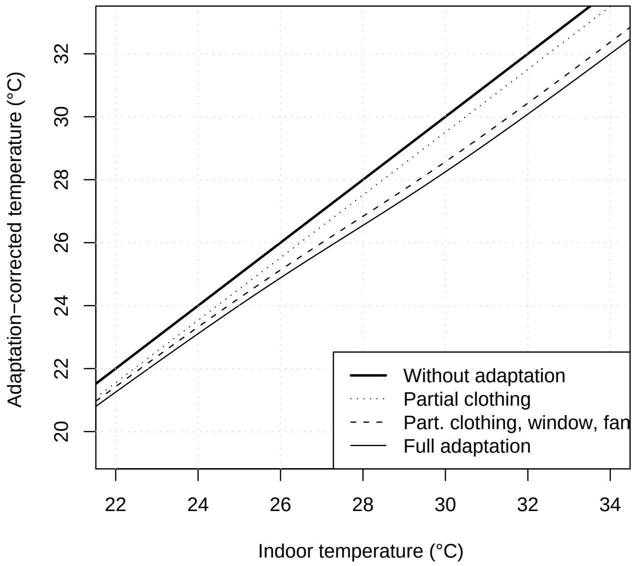


Figure 6.22: Adaptation-corrected versus actual indoor temperature for four different cases of available adaptive opportunities

corrected temperature θ_{ad}:

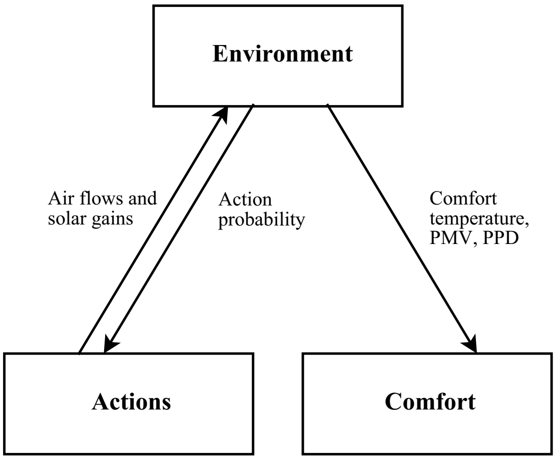
$$\theta_{\text{ad}} = \theta_{\text{in}} - \tilde{\theta}(\theta_{\text{in}}) = \theta_{\text{in}} - \sum_i p(\theta_{\text{in}}) \cdot \beta_i, \tag{6.18}$$

where $\tilde{\theta}(\theta_{\text{in}})$ is the adaptive correction defined from the empirical adaptive increments β_i and $p(\theta_{\text{in}})$, the probabilities for the controls to be used.

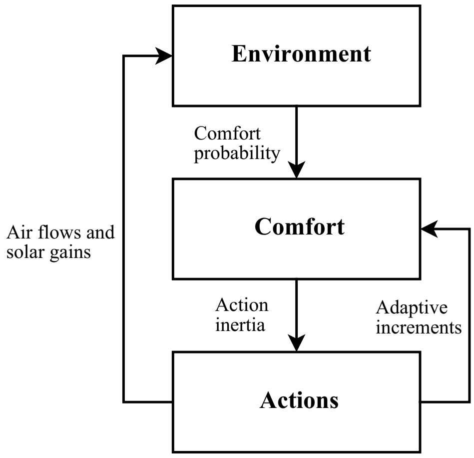
With this definition, we can study the variation of θ_{ad} (the perceived temperature accounting for adaptive actions) as θ_{in} increases. We observe in Figure 6.22 for example that the difference between θ_{ad} and θ_{in} increases for higher temperatures, which corresponds to a higher probability of performing adaptive actions[11] and therefore to a benefit from the associated adaptive increments. It is worth noting that some actions considered for the calculation of θ_{ad} (like the use of a fan or changing the clothing level) alter the temperature at which one is neutral, without altering the actual temperature; while others, such as opening a window, chiefly alter the room temperature itself. For this latter case, the deduced increment already accounts both for this physical effect and the non-physical effects dealt with in the previous case.

The four presented curves correspond to different available adaptive opportunities. In the case of no possible adaptation, θ_{in} and θ_{ad} coincide. When actions on windows, fans and clothing are possible, the difference between adaptation-corrected and actual temperature becomes significant; for instance when $\theta_{\text{in}} = 28°\text{C}$ the occupant perceives a temperature θ_{ad} of 26.5°C.

[11]The probability for fans to be switched on is deduced from the expression published by Haldi and Robinson [74]. The derivation of ordinal probabilities for clothing levels is provided in Appendix A.



(a) Current view



(b) Proposed view

Figure 6.23: A proposal for a new perspective to understand the interactions between the environment, occupant comfort and adaptive actions

This estimation of the effect of control actions on thermal comfort enables a more rigorous estimation of the acceptable conditions in naturally-ventilated buildings. Furthermore this analysis underlines the positive impact of unconstrained adaptation on occupant comfort and on target indoor temperatures, which can be set higher or lower provided that occupants have adequate freedom to use windows, fans and to adapt their clothing levels.

This quantification of the effects of adaptive actions on occupants' comfort partly confirms our preliminary hypothesis (Section 1.2). However, the separation of physical, physiological and psychological feedback in this effect remains to be studied.

Finally, our data preclude an examination of the impact of performed actions on subsequent actions (the probability of which may depend on feedback from prior actions); but based on our results, we suggest that θ_{ad} accounts for this effect, and that subsequent action probabilities should take θ_{ad} instead of θ_{in} as an input variable to account for the feedback from prior actions. This assertion needs confirmation following further detailed measurements. Our analysis of actions on windows and shading devices did not isolate any significant interaction between these latter, however various adaptive actions linked to thermal comfort (windows, fans, clothing) are expected to be connected.

6.7 Discussion and perspectives

We have studied the conditions which lead to the states of thermal and visual comfort in office buildings and the impact of adaptive opportunities in the thermal case (formulated as adaptive increments). Finally, we have proposed a rigorous approach to examine the link between actions and comfort. We have also explored the interactions between thermal, visual and olfactory comfort (Appendix B).

The above concepts have been integrated into a modelling approach in which we consider interactions between environmental stimuli, occupants' actions and comfort. More specifically, while the classical scheme separately considers the prediction of actions and the assessment of comfort (Figure 6.23(a)), our findings allow us to consider comfort and actions as a single concept through the notion of inertia, and to assess the impact of actions on comfort using adaptive increments (Figure 6.23(b)).

These adaptive increments reformulate the impact of physical variables (such as air speed) modified by occupants' adaptive actions, together with possible psychological factors linked with the ability to act on the environment. They provide therefore a useful shortcut to detailed simulations involving dynamic human body models [129, 130, 131, 132] coupled with CFD simulation [133, 134, 135], while explicitly considering occupant adaptation to his environment.

However, improving computing capabilities open new perspectives in the longer term to such detailed simulation of the interactions between human comfort and the environment, provided that the crucial mechanisms of adaptation are considered. But for the time being the approach developed in this chapter offers a good compromise between simplified adaptive models and intractable complexity.

Chapter 7

Integration with building simulation software

Livet forstås baglæns, men må leves forlæns.
Life can only be understood backwards but it must be lived forward.
Søren Aabye Kierkegaard (1813-1855)
Journalen JJ:167 (1843)

This chapter focuses on the integration and application of our models with dynamic building and urban simulation tools. The procedures for model integration are described in detail in Section 7.1, where we also discuss the integration of models of other actions of interest. The specific situation of simulation at the urban scale is discussed in Section 7.2. We conclude in Section 7.3 by discussing a proposition for behavioural modelling using an agent-based approach.

7.1 Implementation in building simulation

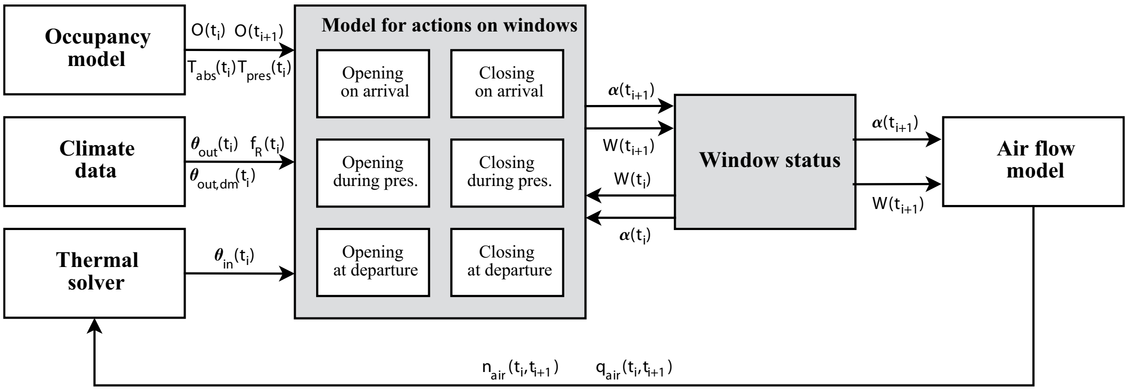
Le simple est toujours faux. Ce qui ne l'est pas est inutilisable.
What is simple is false, what is not is unusable.
Paul Valéry (1871-1945), Œuvres II (1942)

We present in this section proposals according to which the stochastic models developed in this thesis should be integrated with dynamic building simulation software. Their source code will shortly be available at the home page of the Sustainable Urban Development Group of the Solar Energy and Building Physics Laboratory [145].

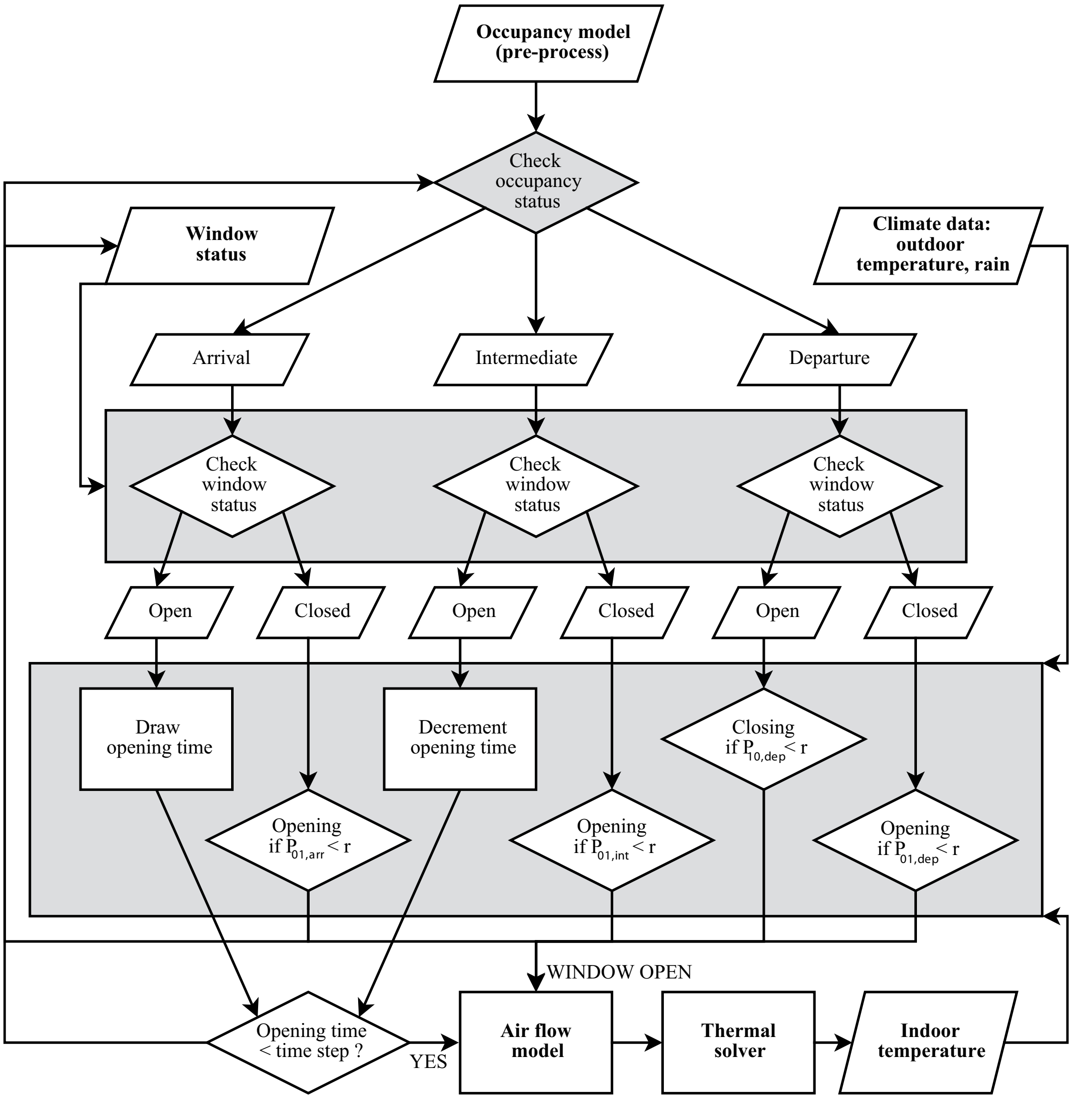
7.1.1 Implementation of the window model

Our model for the prediction of actions on windows is more complex than previously published approaches. Its implementation however is relatively straightforward (Figure 7.1).

Its inputs include the previous, current and next occupancy states with their associated durations, as well as the outdoor thermal conditions and rain presence. The former may be calculated in a pre-process using a model such as that due to Page et al. [28, 29] whereas the latter are readily available from climate files. The principal difficulty lies in



(a) General scheme of the inputs and outputs of the model for actions on windows



(b) Detailed scheme for the implementation of the model for actions on windows

Figure 7.1: Integration of the model for actions on windows in building simulation tools

the dynamic prediction of indoor temperature, resulting from previous actions on windows and the associated impact of their induced air flows. However many detailed dynamic thermal simulation programs such as ESP-r are coupled with bulk airflow models which resolve for both wind and buoyancy pressures.

Some possibly influential factors could not be studied with our data. These include window opening angles α and indoor pollutant concentration, whose dynamic prediction needs a specific sub-model (not displayed here) taking for example occupancy, ventilation and material off-gassing as inputs.

We present here the detailed steps for the implementation of our discrete-time Markov process hybridised with a Weibull distribution to predict opening durations, retained in Chapter 4 as the most accurate approach. Action probabilities are formulated as:

$$P_{01,\text{arr}} = \frac{\exp(-13.88 + 0.312\theta_{\text{in}} + 0.0433\theta_{\text{out}} + 1.862 f_{\text{abs,prev}} - 0.45 f_R)}{1 + \exp(-13.88 + 0.312\theta_{\text{in}} + 0.0433\theta_{\text{out}} + 1.862 f_{\text{abs,prev}} - 0.45 f_R)}, \tag{7.1}$$

$$P_{10,\text{arr}} = \frac{\exp(3.97 - 0.286\theta_{\text{in}} - 0.0505\theta_{\text{out}})}{1 + \exp(3.97 - 0.286\theta_{\text{in}} - 0.0505\theta_{\text{out}})}, \tag{7.2}$$

$$P_{01,\text{int}} = \frac{\exp(-12.23 + 0.281\theta_{\text{in}} + 0.0271\theta_{\text{out}} - 8.78 \cdot 10^{-4} T_{\text{pres}} - 0.336 f_R)}{1 + \exp(-12.23 + 0.281\theta_{\text{in}} + 0.0271\theta_{\text{out}} - 8.78 \cdot 10^{-4} T_{\text{pres}} - 0.336 f_R)}, \tag{7.3}$$

$$P_{01,\text{dep}} = \frac{\exp(-8.75 + 0.1371\theta_{\text{out,dm}} + 0.84 f_{\text{abs,next}} + 0.83 f_{GF})}{1 + \exp(-8.75 + 0.1371\theta_{\text{out,dm}} + 0.84 f_{\text{abs,next}} + 0.83 f_{GF})}, \tag{7.4}$$

$$P_{10,\text{dep}} = \frac{\exp(-8.54 + 0.213\theta_{\text{in}} - 0.0911\theta_{\text{out,dm}} + 1.614 f_{\text{abs,next}} - 0.923 f_{GF})}{1 + \exp(-8.54 + 0.213\theta_{\text{in}} - 0.0911\theta_{\text{out,dm}} + 1.614 f_{\text{abs,next}} - 0.923 f_{GF})}, \tag{7.5}$$

and the density of the probability distribution of opening durations as:

$$f_{\text{op}}(t) = \lambda\alpha(\lambda t)^{\alpha-1}\exp(-(\lambda t)^{\alpha}), \quad \text{with} \quad \alpha = 0.418, \lambda = 1/\exp(2.151 + 0.1720\theta_{\text{out}}). \tag{7.6}$$

We assume that climate data are available and that occupancy is first predicted through a pre-processor for the whole simulation period. A detailed scheme describing the implementation procedure for a dynamic simulation of window states $W(t)$ for time steps of length $\delta t = (t_{i+1} - t_i)$ is provided in Figure 7.1(b), which involves the following steps:

1. The occupancy status (absence, arrival, ongoing presence or departure) is retrieved, together with the concomitant presence or absence durations. The variables $\theta_{\text{out}}(t_i)$, $\theta_{\text{out,dm}}(t_i)$ and $f_R(t_i)$ are obtained from the climate data and $\theta_{\text{in}}(t_i)$ from a coupled thermal solver.

2. Case 1. If the occupant is absent, the window state is set as identical to its previous state: $W(t_{i+1}) = W(t_i)$.

2. Case 2. If the occupant arrives and the window is closed ($W(t_i) = 0$):

- **a)** We compute $p(t_i) = P_{01,\text{arr}}(t_i)$ and draw a random number $0 \leq r_i \leq 1$ from a uniform distribution.
- **b1)** If $p(t_i) > r_i$ we set $W(t_{i+1}) = 0$.
- **b2)** If $p(t_i) \leq r_i$, we draw a random number d_i from the Weibull distribution $f_o(\theta_{\text{out}}(t_i))$. If $d_i < \delta t$, we set $W(t_{i+1}) = 0$. If $d_i \geq \delta t$, we set $W(t_{i+1}) = 1$ and $d_{i+1} = d_i - \delta t$.

2. Case 3. If the occupant arrives and the window is open ($W(t_i) = 1$), we draw a random number d_i from the Weibull distribution $f_{\text{op}}(\theta_{\text{out}}(t_i))$, and perform the same procedure as in Case 2 (b2).

2. Case 4. If the occupant is in intermediate presence and the window is closed ($W(t_i) = 0$):

1. We compute $P_{01,\text{int}}(t_i)$ and draw a random number $0 \le r_i \le 1$ from a uniform distribution.
2. If $p(t_i) > r_i$ we set $W(t_{i+1}) = 0$.
3. If $p(t_i) \le r_i$, we draw a random number d_i from the Weibull distribution $f_{\text{op}}(\theta_{\text{out}}(t_i))$, and perform the same procedure as in Case 2 (b2).

2. Case 5. If the occupant is in intermediate presence and the window is open ($W(t_i) = 1$), we retrieve the decremented opening duration d_i calculated at the previous opening. If $d_i < \delta t$, we set $W(t_{i+1}) = 0$. If $d_i \ge \delta t$, we set $W(t_{i+1}) = 1$ and $d_{i+1} = d_i - \delta t$.

2. Case 6. If the occupant leaves:

a) We compute $p(t_i) = P_{01,\text{dep}}(t_i)$ if $W(t_i) = 0$ (or $p(t_i) = P_{10,\text{dep}}(t_i)$ if $W(t_i) = 1$) and draw a random number $0 \le r_i \le 1$ from a uniform distribution.

b1) If $p(t_i) > r_i$ we set $W(t_{i+1}) = 0$ (or $W(t_{i+1}) = 1$).

b2) If $p(t_i) \le r_i$, we set $W(t_{i+1}) = 1$ (or $W(t_{i+1}) = 0$).

3. The volume of air exchanged $n_{\text{air}}(t_i, t_{i+1})$ is computed, the thermal solver predicts $\theta_{\text{in}}(t_{i+1})$ and we start the next time step.

7.1.2 Implementation of the blinds model

The model for the prediction of actions on blinds (Figure 7.2) also takes pre-processed occupancy states as an input as well as outdoor illuminance which is readily derived from diffuse horizontal illuminance using an appropriate luminous efficacy model. However, at each step a prediction of indoor illuminance is also required, so that our dynamic thermal model should also be coupled with a daylight model.

Based on the analyses of Chapter 5, the model for the prediction of actions on lower blinds consists of the following action probabilities:

$$P_{\text{lower,arr}} = \frac{\exp(-7.41 + 10.35 \cdot 10^{-4} E_{\text{in}} + 2.17 B_L)}{1 + \exp(-7.41 + 10.35 \cdot 10^{-4} E_{\text{in}} + 2.17 B_L)}, \tag{7.7}$$

$$P_{\text{raise,arr}} = \frac{\exp(-1.520 - 6.54 \cdot 10^{-4} E_{\text{in}} - 3.139 B_L)}{1 + \exp(-1.520 - 6.54 \cdot 10^{-4} E_{\text{in}} - 3.139 B_L)}, \tag{7.8}$$

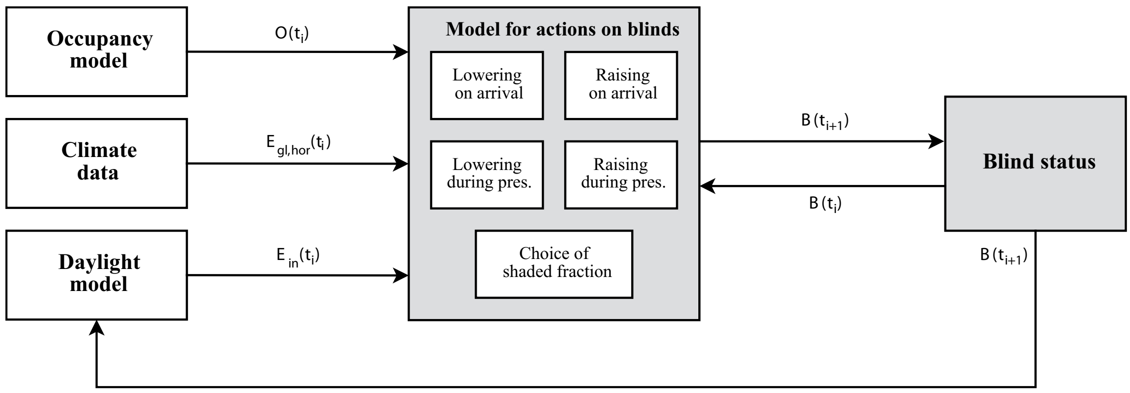
$$P_{\text{lower,int}} = \frac{\exp(-8.013 + 8.41 \cdot 10^{-4} E_{\text{in}} + 1.270 B_L)}{1 + \exp(-8.013 + 8.41 \cdot 10^{-4} E_{\text{in}} + 1.270 B_L)}, \tag{7.9}$$

$$P_{\text{raise,int}} = \frac{\exp(-3.625 - 2.76 \cdot 10^{-4} E_{\text{in}} - 2.683 B_L)}{1 + \exp(-3.625 - 2.76 \cdot 10^{-4} E_{\text{in}} - 2.683 B_L)}. \tag{7.10}$$

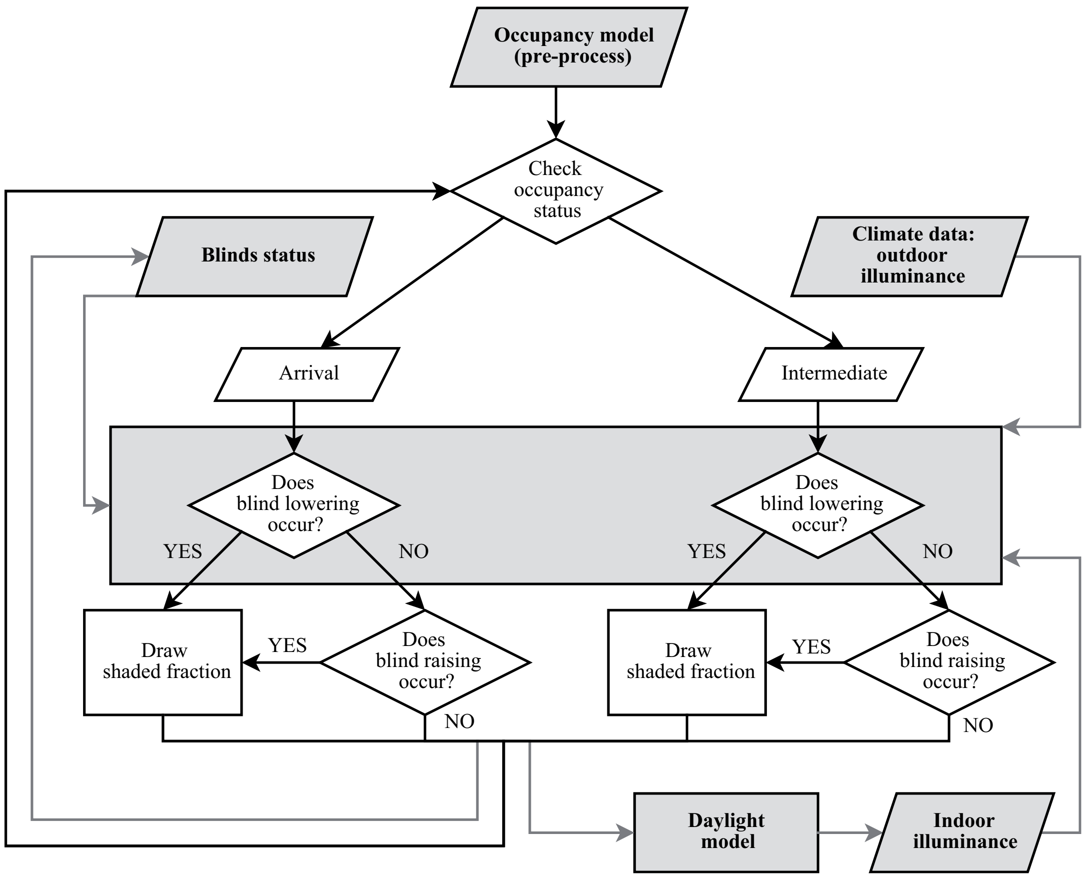
If an action is predicted, the probabilities to perform it up to a full (un)shaded fraction are:

$$P_{\text{full lower}} = \frac{\exp(-0.27 + 0.91 \cdot 10^{-6} E_{\text{gl,hor}} - 2.23 B_L)}{1 + \exp(-0.27 + 0.91 \cdot 10^{-6} E_{\text{gl,hor}} - 2.23 B_L)}, \tag{7.11}$$

$$P_{\text{full raise}} = \frac{\exp(0.435 - 2.31 \cdot 10^{-5} E_{\text{gl,hor}} + 1.95 B_L)}{1 + \exp(0.435 - 2.31 \cdot 10^{-5} E_{\text{gl,hor}} + 1.95 B_L)}. \tag{7.12}$$



(a) General scheme of the inputs and outputs of the model for actions on blinds



(b) Detailed scheme for the implementation of the model for actions on blinds

Figure 7.2: Integration of the model for actions on blinds in building simulation tools

Finally, if a partial action is predicted the increase in shading ΔB is drawn from the distribution

$$f(\Delta B) = \frac{\alpha}{\lambda}(\frac{\Delta B}{\lambda})^{\alpha-1}\exp(-(\frac{\Delta B}{\lambda})^{\alpha}), \quad \text{with} \quad \alpha = 1.708, \lambda = \exp(-2.294+1.522\cdot B_{\mathrm{L,init}}). \tag{7.13}$$

In the case of actions on upper blinds, we have:

$$P_{\mathrm{lower,arr}} = \frac{\exp(-7.29 + 9.48\cdot 10^{-4}E_{\mathrm{in}} + 2.18B_U + 6.66\cdot 10^{-6}E_{\mathrm{gl,hor}})}{1+\exp(-7.29 + 9.48\cdot 10^{-4}E_{\mathrm{in}} + 2.18B_U + 6.66\cdot 10^{-6}E_{\mathrm{gl,hor}})}, \tag{7.14}$$

$$P_{\mathrm{raise,arr}} = \frac{\exp(-1.699 - 5.24\cdot 10^{-4}E_{\mathrm{in}} - 3.916B_U - 21.8\cdot 10^{-6}E_{\mathrm{gl,hor}})}{1+\exp(-1.699 - 5.24\cdot 10^{-4}E_{\mathrm{in}} - 3.916B_U - 21.8\cdot 10^{-6}E_{\mathrm{gl,hor}})}, \tag{7.15}$$

$$P_{\mathrm{lower,int}} = \frac{\exp(-8.211 + 8.34\cdot 10^{-4}E_{\mathrm{in}} + 1.533B_U + 5.69\cdot 10^{-6}E_{\mathrm{gl,hor}})}{1+\exp(-8.211 + 8.34\cdot 10^{-4}E_{\mathrm{in}} + 1.533B_U + 5.69\cdot 10^{-6}E_{\mathrm{gl,hor}})}, \tag{7.16}$$

$$P_{\mathrm{raise,int}} = \frac{\exp(-3.629 - 2.90\cdot 10^{-4}E_{\mathrm{in}} - 3.365B_U - 16.86\cdot 10^{-6}E_{\mathrm{gl,hor}})}{1+\exp(-3.629 - 2.90\cdot 10^{-4}E_{\mathrm{in}} - 3.365B_U - 16.86\cdot 10^{-6}E_{\mathrm{gl,hor}})}, \tag{7.17}$$

$$P_{\mathrm{full\ lower}} = \frac{\exp(-0.435 + 2.50\cdot 10^{-6}E_{\mathrm{gl,hor}} + 0.150B_U)}{1+\exp(-0.435 + 2.50\cdot 10^{-6}E_{\mathrm{gl,hor}} + 0.150B_U)}, \tag{7.18}$$

$$P_{\mathrm{full\ raise}} = \frac{\exp(1.543 - 2.12\cdot 10^{-5}E_{\mathrm{gl,hor}} - 0.56B_U)}{1+\exp(1.543 - 2.12\cdot 10^{-5}E_{\mathrm{gl,hor}} - 0.56B_U)}. \tag{7.19}$$

This model has thus a similar level of complexity to that of the model for windows, except that its internal structure includes two sub-models: one for the prediction of actions and another for the choice of shaded fraction.

Based on our results, we have developed a simple algorithm for the simulation of blind usage, which is based on the following steps (see Figure 7.2(b)):

- The occupancy status is checked. If no occupant is present, the state of blinds remains constant.
- If $P_{\mathrm{lower}} \geq P_{\mathrm{raise}}$, we use the Monte-Carlo method on P_{lower} to determine whether the blind is to be lowered. If no lowering is predicted, the Monte-Carlo method determines whether a raising action occurs through P_{raise}. The order of this procedure is inverted if $P_{\mathrm{lower}} < P_{\mathrm{raise}}$.
- If an action is predicted, the Monte-Carlo method applied on $P_{\mathrm{full\ lower}}$ or $P_{\mathrm{full\ raise}}$ predicts whether the blind is set up to a full extent. Otherwise the new shaded fraction is drawn from a Weibull distribution.
- This procedure is performed sequentially for each type of blind.

7.1.3 Integration of complementary stochastic models

The relevant non-deterministic behavioural processes discussed in Chapter 1 are related in a complex scheme presented in Figure 7.3, where the central model first predicts occupants' presence. In this section, we describe the features of the other existing or future models influencing energy flows in buildings.

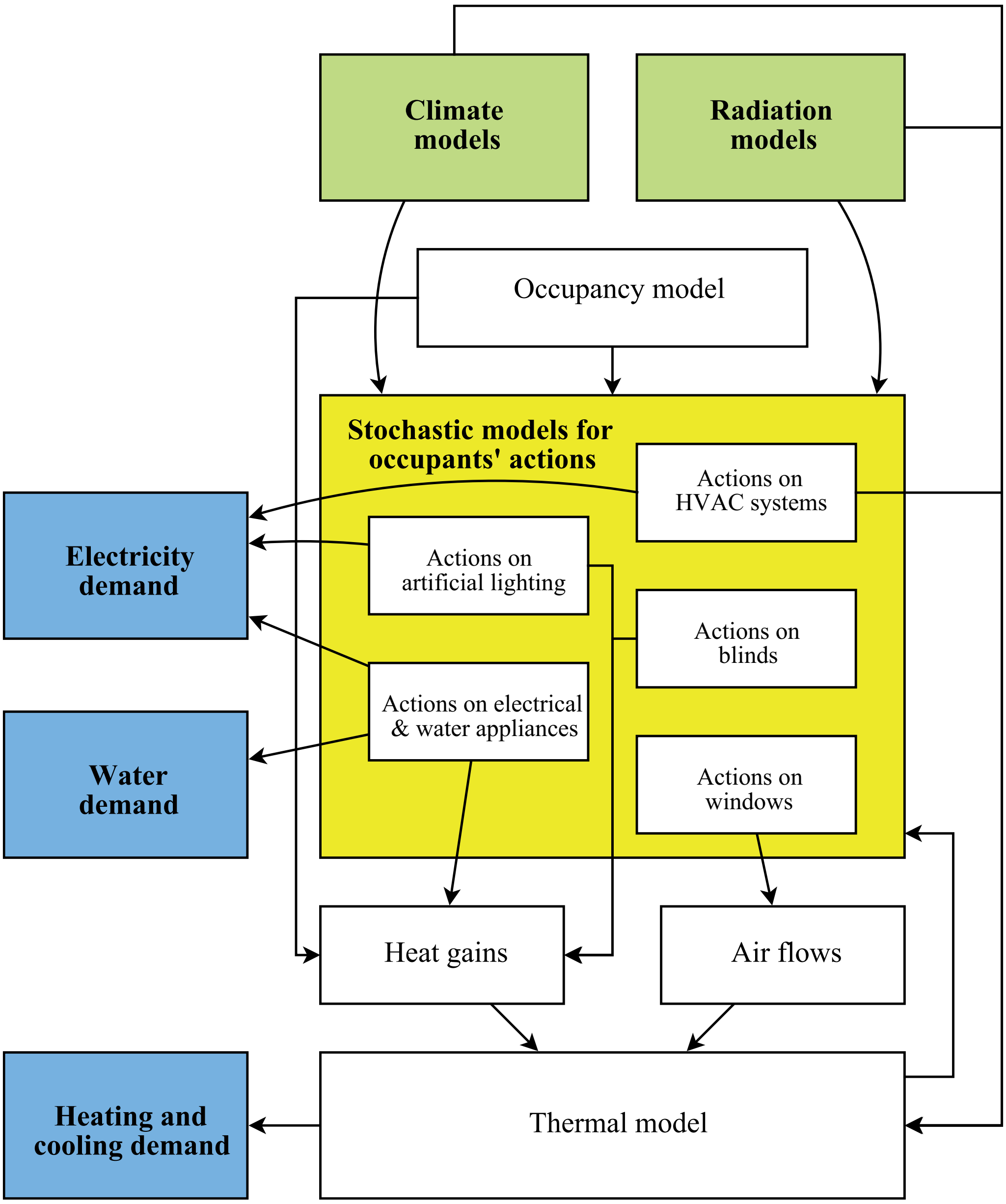


Figure 7.3: The set of relevant stochastic models for occupants' actions in dynamic building thermal simulation and their interactions

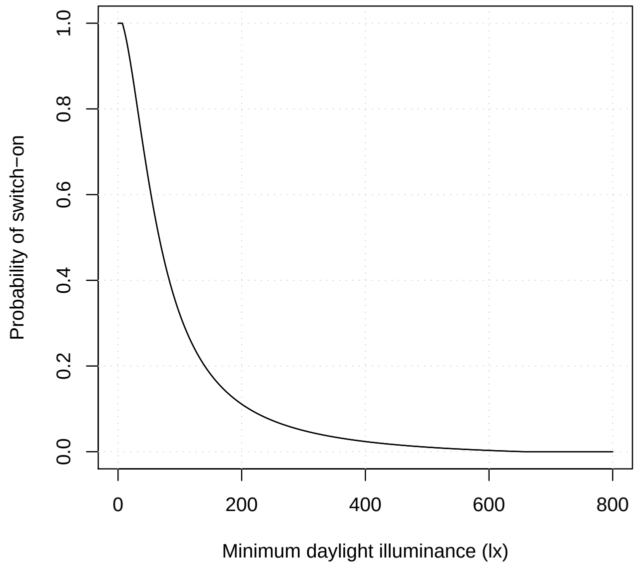


Figure 7.4: Hunt's function for the probability of switch-on on arrival

Occupancy

As suggested in the preceding sections occupants' interactions depend on their presence, so that a model of presence is central to our family of stochastic behavioural models.

A sound approach for the modelling of occupancy was developed by Page et al. [28, 29], based on an inhomogeneous Markov chain, assuming the independence of occupants and the superposition of two modes of absence: long absences and daily movements. The inputs are a time-varying profile of the probability of presence and a parameter of mobility (the ratio between the probabilities of unchanged and changed presence).

Artificial lighting

The initial research performed in this domain was conducted by Hunt in 1979 [146, 147, 148], who provided a description of the probability of people switching on the lights on arrival as a function of the minimum daylight illuminance $E_{\min}$ in the working area[1] (Figure 7.4):

$$p = \begin{cases} 1 & (\text{if} \quad \log_{10}(E_{\min}) \leq 0.843) \\ a + c/\left(1 + \exp(-b(\log_{10}(E_{\min} - m)))\right) & (\text{if} \quad 0.843 < \log_{10}(E_{\min}) < 2.818) \\ 0 & (\text{if} \quad \log_{10}(E_{\min}) \geq 2.818) \end{cases}$$

with $a = -0.0175$, $b = -4.0835$, $c = 1.0361$, $m = 1.8223$ (lx).

This attempt was seen as a basis for more accurate predictions of the energy consumed in buildings, and represents the first stochastic model of occupants' behaviour; it has been

[1] This latter variable was retained among other visual variables as it was observed to be the best predictor.

integrated with ESP-r and SERI-RES [149] in the 1980s. However, the limited scope of this model called for further developments. Consequently Pigg produced a model to predict the switching off of lights based on foreseen absence duration [150] and Reinhart worked on the probability of switch-on during the day (intermediate switch-on) [79, 151]. This resulted in the model proposed in Lightswitch-2002.

Electrical and water appliances

We may distinguish two main approaches for the simulation of the use of electrical appliances: *bottom-up models* which perform explicit simulation of the use of individual appliances resulting in total load through aggregation of their predictions; and *top-down models* that treat the appliances to simulate as a black box energy sink based on historic data to project future results.

This latter approach, used for instance by Stokes [152] and Steemers and Yun [153], allows for example for the identification of significant factors influencing the yearly energy consumption of households.

Bottom-up approaches proposed by Paatero [154] and later by Page [28] are promising but need more solid calibration and validation tests. Page distinguishes appliances by their operation mode, which results in distinction between cases where they are (a) neither switched on nor off by occupants (eg. fridge), (b) switched on by an occupant but switched off independently of occupant presence (eg. washing machine) and (c) switched on and off by occupants (eg. television). Occupant behaviour (and its variability) can be explicitly considered and simulations of appliance use are performed with respect to their switching mode. In principle, this results in detailed predictions of power demand. Finally, with this approach the consumption of water by appliances and the associated rejection of waste water can also be handled, by extension.

Heating, ventilation and air-conditioning

Research is lacking in the study of human interactions with HVAC devices. We may however hypothesise that a Markov process based on local indoor thermal stimuli would provide a valid approach. Some research in this direction was performed by Tanimoto and Hagishima [155] on the use of air-conditioning devices in dwellings, where probabilities of switch-on and switch-off are proposed as logistic functions of air temperature; and also by Schweiker and Shukuya [156] on the use of air-conditioning during nighttime. However, further experimental data are required to validate these hypotheses and to reach a sufficient level of generality.

7.2 Implementation in urban simulation

The initial motivation for the work described in this thesis was to account for the impact of occupants' behavioural diversity on simulations of urban resource (energy and matter) flows; more specifically to develop and integrate a comprehensive family of stochastic models (which may optionally be substituted by the used deterministic profiles and rules) of occupants' presence and behaviour within CitySim [157] which is a software currently under development.

CitySim's main purpose is to simulate building-related energy demand, storage and supply at the urban scale, where it is particularly important to achieve a good compromise

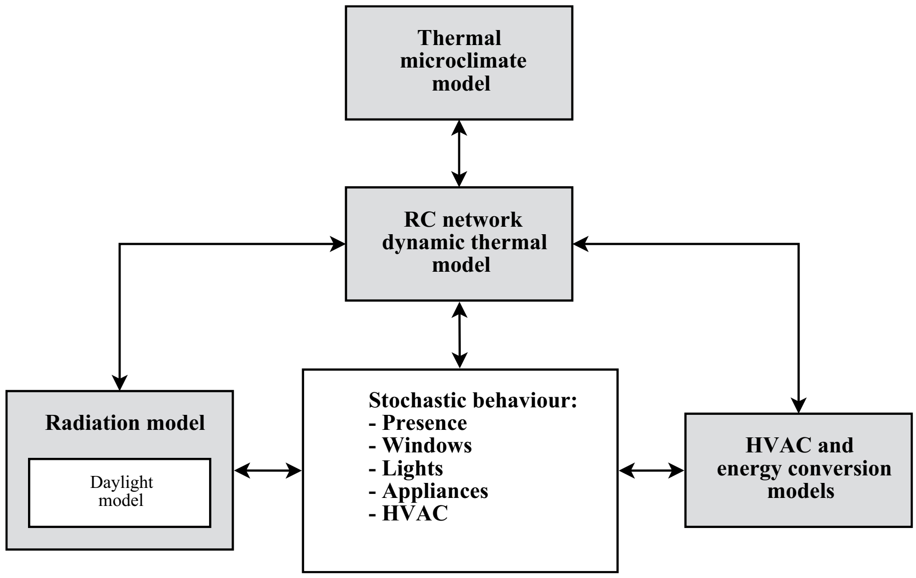


Figure 7.5: General scheme of the structure of CitySim

between modelling accuracy and computational effort. Its general structure, shown in Figure 7.5, currently includes the following core models:

- **Thermal solver.** CitySim uses a thermal model developed by Kaempf and Robinson [158], based on the analogy with an electrical circuit represented by a resistor-capacitor network.

- **Radiation model.** The Simplified Radiosity Algorithm (SRA) of Robinson and Stone [159, 160] is used to solve for the shortwave and longwave irradiance incident on the surfaces defining our urban scene, as well as their luminance. This latter may also be used to solve for indoor illuminance.

- **Plant and equipment models.** An HVAC model computes the psychrometric state (enthalpy) of the air at each stage in its supply (eg. outside, heat recovered, cooled and de-humidified, re-heated, supply). The family of models for energy conversion systems accounts for a range of technologies that provide or store heat and electricity to buildings. These include a thermal storage tank model for hot and cold fluids, boilers, heat pumps, cogeneration systems, combined cogeneration and heat pump systems, solar thermal collectors, photovoltaic cells and finally wind turbines. These models are discussed in detail by Kaempf [161]

Finally, the programming structure of CitySim is designed to support in the future an agent-based simulation (Section 7.3) of occupants' presence and behaviour at the urban scale. Work on their integration will start imminently.

Apart from the inclusion of behavioural models, a future planned development is the inclusion of the simulation of the transport of goods and people – a key resources consumer in the urban environment. For this we plan to couple CitySim with the Multi-Agent Transport Simulation Toolkit (MATSim-T) [162], which simulates the sub-hourly transport of individual people within a given urban scene, based on the transport network nodes and the links between them as well as the locations of activities.

7.3 Towards an agent-based modelling paradigm

I, at any rate, am convinced that He does not throw dice.
Albert Einstein (1879-1955), Letter to Max Born (1926)

Throughout our development of models for occupants' actions we have consistently focused upon individuals acting under the influence of stimuli. *Agent-based models* (ABM), in which the decisions of individual agents are simulated in order to observe their aggregated effects on the whole system, provide a natural framework for modelling individuals' actions in response to representations of their personal preferences.

7.3.1 Simulation steps

Population and characteristics. A simulation would start with the creation of a population of a determined number of agents. To enable the modelling of the diversity among them, behavioural attributes would be stochastically assigned to each agent, through the definition of characteristics of interest for the modelling context. To ensure a correct representation of the system to be modelled, these attributes must be drawn from suitable multivariate distributions based on census and survey data. Characteristics of potential interest are listed below (Table 7.1), some of which may ultimately prove to be inter-related with complex correlations.

Sociology	**Behaviour**
Age, sex	Occupancy profiles
Marital status	Mobility parameters
Relation with other agents	Available activities
Income, wealth	Preferred activities
Disability	Activity profile
Car, motorbike ownership	Thermal and visual comfort prob.
...	Action inertia
...	...

Table 7.1: Hypothetical list of agents' attributes of interest

This generalist selection process is also flexible; it can be easily adapted to different contexts or to future changes in society (eg. transportation improvements, population aging, land use changes, etc), based on new data, to reliably predict the impact of any scenario of interest.

Finally, this approach is in agreement with subject-matter considerations in building simulation: occupants' behaviour is a process generated by the intersection of a specific occupant with a specific building. Agent-based models may explicitly describe this interaction through the simulation of a specific agent acting in a given zone.

Presence. The next step is to predict the presence of agents in the simulated system. To do this, a model with generic input parameters such as that developed by Page (Section 7.1.3) is appropriate; as the occupancy profile and the parameter of mobility may be adjusted to distinct behaviours and schedules. The scope of this model needs however to be extended in this context, as in its basic form it predicts a binary response which

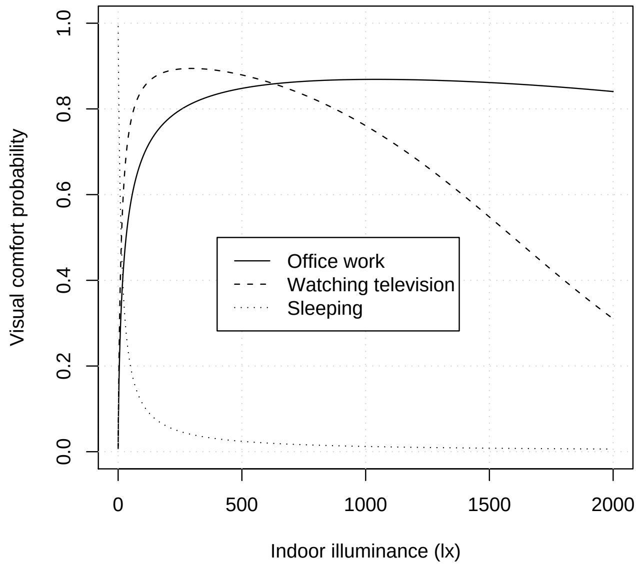


Figure 7.6: Hypothetical visual comfort probabilities with respect to occupant activity

is absence or presence (so that agents simply appear and disappear). The model output needs thus to be extended to continuously track agents' locations.

Alternatively and as noted in Section 7.2, a third party transport simulation program may be used to simulate occupants' arrival and departure times from buildings with our simulation scene. But again these agents' locations within buildings would then need to be tracked.

Events. Occupants' activities are events which are potentially of interest in the simulation. We must bear in mind that the data used in this research work come from office environments exclusively, where activities are usually limited to reading, writing and computer based tasks. This specific situation can have a strong impact on actions and comfort and lead to rather specific results. For instance, the probability of visual comfort is obviously influenced by the type of activity performed (Figure 7.6) and therefore indirectly impacts the use of controls of interest such as lighting and blinds. Furthermore, predictions of electrical appliance use may be seen as a post-processor of a model for occupants' activities.

It is thus of interest to explicitly model activities, particularly in the context of residential buildings, from which comfort and action probabilities would be deduced. To do this, very limited published experimental evidence exists (for instance the studies of Tanimoto et al. [163, 164] and Tabak and de Vries [165]), but the analysis of data from time-use surveys is a promising approach, which could possibly be implemented as a pre-processor of actions, in conjunction with the occupancy model. Thus, from our original scheme considering presence as the sole input for actions we propose to head towards the agent-based structure presented in Figure 7.7.

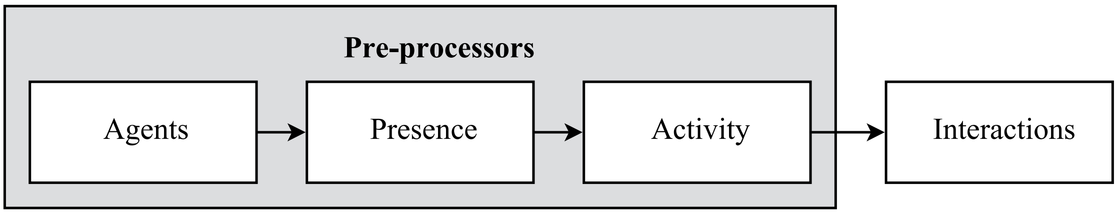


Figure 7.7: From agents to interactions, through occupant presence and activity

Interactions. Once occupants' locations are determined, their behaviour towards actions of interest may be modelled. The models that we developed for actions on windows and blinds are directly applicable to an agent-based modelling paradigm, as we explicitly considered actions performed by individuals and assessed the associated behavioural diversity. Similarly, the model Lightswitch-2002 [79, 151] for the simulation of electric lighting use is appropriate for this purpose.

For the prediction of appliance use a bottom-up approach such as that proposed by Page [28] or Paatero [154] is necessary, as top-down approaches do not explicitly consider electricity use through specific actions. But as mentioned earlier such models need further calibration and validation. Nevertheless this approach is more informative and flexible: the distribution of electricity needs is readily obtained and changes in technology or household equipment are easily integrated through the input parameters.

Finally, a wide calibration basis including data from different types of buildings, climates and environment will strengthen the scope of application of the models. Further acquisition of data is therefore to be envisaged.

7.3.2 Modelling of agents' diversity

Summarising individual patterns. The models for the probabilities of actions on windows and shading devices developed in Chapters 4 and 5 are all formulated as logistic models. Each probability depends thus on two parameters only (an intercept a and a slope b) which completely determine the model. Therefore, the lists of obtained regression parameters (Tables 4.8 and 5.5) offer comprehensive information on individual action probabilities.

The intercept is a location parameter: its purpose is to express the position of the logistic curve with respect to the driving variables. As mentioned earlier, in the case of univariate models the characteristic value $x_{50} = -a/b$ offers direct interpretability in terms of values of the driving stimulus associated with a 0.5 action probability. Based on the values of $\theta_{\mathrm{in},50}$ and $E_{\mathrm{in},50}$ derived for individual occupants, we may thus operate a first classification of their sensitivity to environmental stimuli, those displaying lower values being more likely to perform actions at low values of θ_{in} and E_{in}.

In the particular case of actions on windows, the frequency of occurrence of predictive behaviour may be used to assign decreasing action probabilities for high values of θ_{out}, based on observed polynomial logistic regression or on non-parametric estimates. Similarly as action probabilities, individual distributions for opening and closing durations are fully determined by the scale and the shape.

Simulating behavioural diversity. The random selection of individuals' behaviours through their regression coefficients can be performed as a further pre-process in the simulation. As noticed in Sections 4.2.5 and 5.6, the regression parameters of separate action probabilities (eg. on arrival and during presence) are correlated (Figures 4.10 and 5.12) and thus cannot be simulated independently.

The sampling of individual behavioural profiles may proceed in two ways. A first simple possibility is to directly use the values displayed in Tables 4.8 and 5.5. For this, if N personal profiles are known, an individual $i \in [1, N]$ may be randomly chosen from a uniform distribution for subsequent simulation using the corresponding regression coefficients. Indeed a new population of individuals may be selected and simulated by repeating this process.

A more refined approach to simulate occupant variability would involve direct sampling from the joint multivariate distribution of regression parameters considering their correlations (graphically summarised in Figures 4.10 and 5.12). If a multivariate normal distribution is found to be unappropriate, the use of methods such as the Hastings-Metropolis Algorithm is necessary (see Ross for further discussion [166]). The advantage of this method over the previous sampling procedure is the generation of new representative behaviours rather than a repetitive use of predetermined regression coefficients. However, a larger dataset is necessary to accurately define the properties of this multivariate distribution.

Chapter 8

Conclusion

Toute autre science est dommageable à celui qui n'a pas la science de la bonté.
All other knowledge is hurtful to him who has not the science of goodness.
Michel de Montaigne (1533-1592), Essays (I.24)

Summary

Based on eight years' continuous measurements of the indoor environment and of the use of its controls, coupled with a specific questionnaire for the assessment of occupant comfort and personal and environmental characteristics, we have performed a careful statistical analysis, which results in the following advances:

- A model for the prediction of actions on windows performed by office occupants, based on a set of rigorously selected explanatory variables, verified by a classical cross-validation procedure and selected among other variants based on a combination of predictive accuracy and computation cost.

- A model to predict office occupants' interactions with shading devices based on a set of rigorously selected explanatory variables (including local visual stimuli) with a formulation that can be adapted to particular types of shading systems without significant structural modifications.

- A probabilistic method for the prediction of human thermal and visual sensation and comfort, that allies simplicity and accuracy.

- A refinement of the accepted adaptive thermal comfort model, which results in a better understanding of the distinct roles of control on the environment, acclimatisation and of individual specificities on variations of comfort temperature.

- A formulation of the link between comfort and actions through the definition of the concept of action inertia, together with evaluation of the feedback from actions on comfort. This closes the loop and places human comfort in its rightful central place.

- A better knowledge of individual specificities towards interactions with building controls and proposed possibilities to account for this diversity.

- A model for predicting clothing insulation whose formulation is in principle easily adaptable to any type of environment. Models of the adaptation of other personal characteristics have also been prepared, including the consumption of drinks and changes in metabolic activity.
- An initial assessment of the interactions between thermal, olfactory and visual comfort.
- An analysis and quantification of the factors influencing perceived productivity in office environments.

Contribution of the proposed stochastic models

The stochstic models for occupants' behaviour open new perspectives for dynamic simulation programs. The following issues currently ignored by deterministic approaches are of particular interest and can be tested in the near future:

- **Increased accuracy.** The integration of occupant behaviour will improve the realism of building simulation results enabling the energy and comfort implications of building design and controls to be more reliably assessed at the design stage.
- **Improved basis for low energy design.** In the particular case of passive and low energy buildings, where the behaviour of occupants has a particularly crucial impact, these methods have a special interest.
- **Study of the variability.** The variability of energy demand and indoor conditions induced by stochastic models can be assessed through repeated simulations, which yields distributions of results rather than deceptive fixed values.
- **Impact of extreme behaviours.** The design of buildings which are robust to a wide range of behavioural types is made possible and verifiable, by directly testing the impact of specific action probabilities on a building's energy balance.

Contribution of the research on comfort

Our research on environmental comfort results in an approach situated at reasonable distance of both the simplicity of currently accepted adaptive thermal comfort models and the precision of detailed thermoregulatory models. Its reasonable level of detail leads to advances in dynamic thermal simulation as well as for developing comfort standards:

- **Conceptual advance.** A general formulation of action probability in terms of comfort probability through action inertia represents a conceptually seductive theoretical unification of concepts usually separately treated.
- **Simplicity and sufficiency.** The proposed formulation of comfort probability as a function with four free parameters is together relatively simple and accurate, where thermal and visual comfort are easily simulated and evaluated through univocal indicators (p_{comf} or p_{discomf}).
- **Adequacy for standards.** The model for the evaluation of comfort temperature represents a significant advance in understanding the weight of driving variables for adaptive comfort. Its relatively simple form may be proposed as a useful complement for revised standards.

- **Generality and versatility.** The revision to the adaptive comfort model is easily applicable to specific buildings with extended or restricted means of available adaptation.
- **Individual and collective satisfaction.** Discomfort probability is a superior index to PPD: the aggregation of individual discomfort probabilities results in a distribution of discomfort probability with respect to diversity in preferences.
- **Comfort for all?** Simultaneously testing individual behaviours and comfort preferences together with environmental responses to these behaviours allows us to determine whether the predicted environmental response (eg. θ_{in}, E_{in}) corresponds well with comfort preferences. Repeating this for the whole population allows us to identify for which there is an environmental mismatch with comfort preferences.

Longer term perspectives

In spite of these advances, there is considerable scope for further improvements in behavioural modelling and its applications in dynamic building and urban simulation tools. For example, the following:

- Testing the increased accuracy of integrating variability among occupants in the simulation, based on the simulation of agents.
- Quantifying the impact of societal changes on energy use; also based on the simulation of agents.
- Agent-based modelling of occupants' behaviour, involving the generation of a population and assignment of their characteristics; the modelling of their transport between buildings and, whilst present, their activities and the possible interactions which depend on these activities.
- This agent-based modelling approach would in the future enable us to model (a) diversity between agents and the energy and comfort implications of this diversity (b) the impact of policy measures to influence occupants' behaviour (eg. purchasing/investment decisions and more environmentally responsible behaviour with respect to resource use)
- Coupling of models of occupants interactions, the indoor microclimatic responses to these actions, the feedback to perceived comfort using a thermoregulatory model and the impact on subsequent interactions.

Appendix A

Adaptation of personal characteristics

The prediction of clothing level, metabolic activity and cold or hot drinks consumption is of indirect relevance to building performance simulation. Unlike windows and blinds, these features do not directly influence the heat gains of a building (although changes in metabolic activity marginally modify occupants' metabolic heat production). However, their impact on occupants' comfort is important (see Chapter 6). Personal characteristics may thus have an indirect influence on other adaptive actions of interest for building energy performance assessment.

This chapter presents our results for the prediction of occupants' choices of clothing level (Section A.1), metabolic activity (Section A.2) and the consumption of cold and hot drinks (Section A.2.3). For each topic we begin with a short review of available previous research.

A.1 Clothing

The issue of clothing can be seen from different perspectives. As pointed out by Morgan and De Dear [167], clothing has an ergonomic function, but it relates also to cultural and social aspects, to personality and corporate identity; while its primary function is to serve as a simple layer of thermal insulation. Although in this work we are particularly interested in the latter function, it is useful to bear in mind other factors which motivate our clothing choices.

A.1.1 State of the art

Clothing has an important impact on the human heat balance. As such the reliable estimation of clothing level is important for the estimation of thermal satisfaction eg. using the ISO-7730 standard [35], where clothing level is a key variable (see Section 6.1). It is however difficult to perform precise estimates from field surveys without invasive methods. Methods thus far employed for the estimation of clothing level are discussed by Olesen [168].

A first field study on clothing behaviour was carried out by Humphreys on secondary school children [169, 170]. It was observed that the proportion of children stripped to the minimum allowed clothing ensemble was significantly correlated with the room temperature. An analytical formulation of this proportion was proposed based on probit analysis.

Although within-day changes of clothing also significantly correlated with room temperature, the amplitude of this pattern was much smaller than day-to-day variations.

Further observations were carried out by Humphreys [171] where the focus was on clothing outdoors. In this a logistic function of temperature for the probability of wearing a light clothing ensemble was proposed. It was found that age and sex were not determinant factors, that humidity was not an important variable and that sunshine had a marginal effect.

Integrating and comparing the observations from these surveys, Humphreys proposed to model the clothing level through an exponentially-weighted moving average of the room temperature [172]. The regression parameters governing the influence of temperature and the exponential decay of the running mean were found to not significantly differ among the datasets.

Measurements performed by Nicol et al. in offices in Pakistan [62, 63] (see Section 4.1.1) showed, using linear regression, that the number of items of clothing worn could be associated with both indoor ($R^2 = 0.653$) and outdoor temperature ($R^2 = 0.666$); likewise with the comfort vote ($R^2 = 0.538$), suggesting that subjects adapt their clothing as a function of perceived thermal comfort. Outside of the interval 20°C to 30°C, clothing insulation was found to remain constant. By way of explanation Nicol et al. proposed that clothing adaptation stops when the limits of acceptable clothing in office environments have been reached.

Observations of clothing changes performed by Newsham and Tiller using a computer-based questionnaire [173, 112] revealed that major or minor clothing adjustments were performed for 15% of the hours preceding the survey. Approximately twice more clothing removals were recorded than clothing additions.

In a detailed field survey of clothing and activity performed on 144 persons, Rowe [174] observed a diurnal change in clothing insulation value for 38% of participants, based on detailed indications of garments worn by respondents.

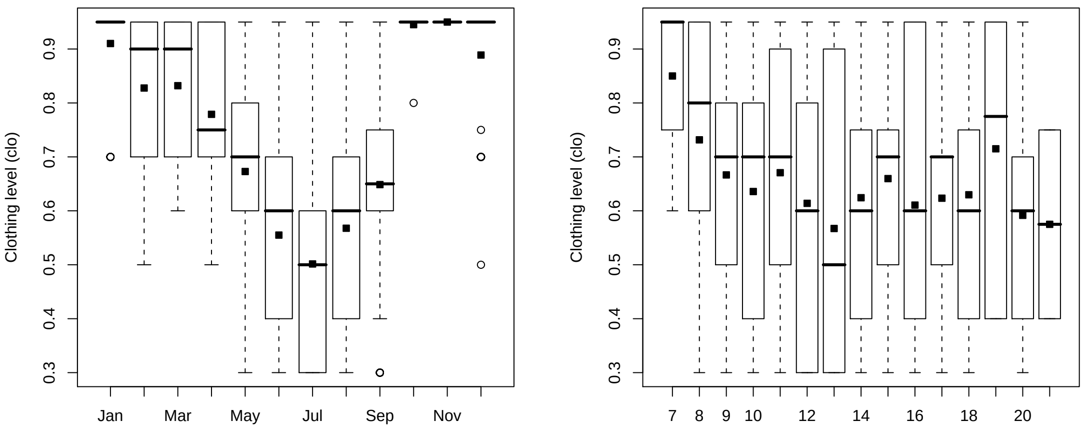
De Dear and Brager [125] observed that there was a significant correlation ($R^2 = 0.25$) between the mean clothing level and indoor temperature. They found an even stronger correlation with daily mean outdoor effective temperature ($R^2 = 0.49$). Morgan and De Dear subsequently summarised these findings, together with a comprehensive literature review completed by additional research [167]. They underline previous experimental evidences that clothing levels worn indoors are affected by gender (women tend to wear less in summer, but more than males in winter), context, corporate dress codes, indoor climate variability and outdoor weather. They go on to present a model to predict the mean daily clothing value as a linear function of the previous mean daily outdoor temperature and the predicted maximum temperature for the current day.

In summary, the use of clothing in offices is well documented by several field studies of variable size and detail. However, improvements are desirable for the following reasons:

- There is a recurrent problem of measurement reliability for clothing level. Often a computer-based electronic questionnaire is used; but self-reported insulation levels are approximate with this method.

- The significant variations due to different dress codes of various office environments are known, but not treated. Furthermore, there is no available research relating to residential environments.

- The influence of indoor or outdoor temperature is well documented. However, data



(a) Monthly evolution

(b) Hourly evolution

Figure A.1: Temporal distribution of observed clothing levels

are often analysed with sub-optimal methods such as linear regression on mean clothing values, with a poor quality of fit, which undermines the utility of the models. A comprehensive, optimal and validated modelling approach therefore remains elusive.

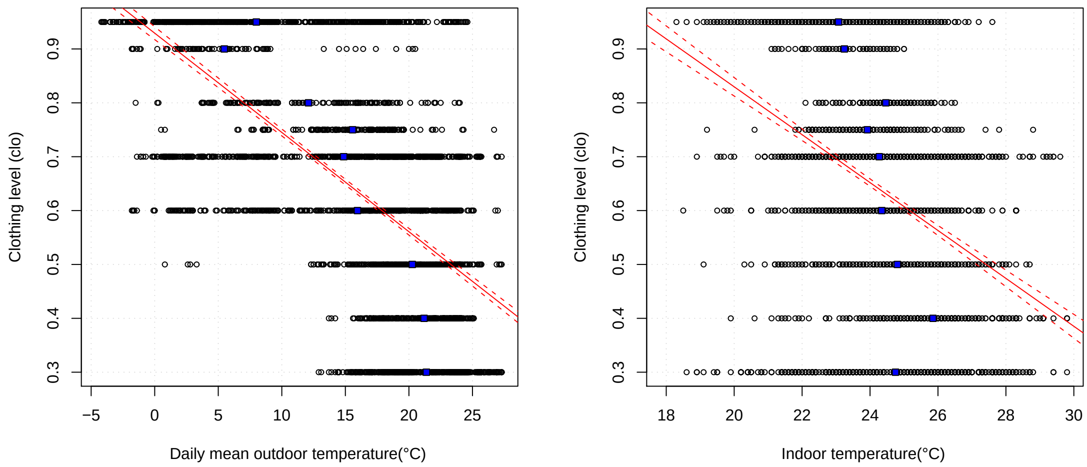
A.1.2 Results

General patterns

A preliminary observation from our dataset shows that occupants rarely adjusted their clothing level during the working day; that they mostly choose a definite set of garments at the start of the day which they do not adjust prior to departure. Among the surveyed periods, lowering clothing level was mentioned for only 1.93% of the preceding hours, while this fraction falls to 0.17% for clothing additions. However, only relatively coarse categories of clothing were incorporated into our questionnaire, so that minor clothing adjustments may have been neglected.

From the above observations it seems necessary to distinguish between adaptation of clothing level between days and the occurrence of such adaptations during the day. By this we mean that occupants may choose their attire at the beginning of the day as a predictive strategy, based on historic experience (e.g. it was warm yesterday and I expect it to be warmer still today, therefore I will reduce my clothing level today) and/or they may wear several layers of clothing and remove these layers as a function of their thermal sensation (e.g. it is cool at the moment, but I expect it to be warm this afternoon, therefore I will provide myself with the possibility to reduce my clothing level during the day). Our preliminary observations indicate that this latter opportunity is seldom exercised which tends to support the approach adopted by Morgan and De Dear [167] to predict a static mean daily clothing level. As an illustration, Figure A.1(a) shows that seasonal variations have a very large amplitude compared to intra-day variations (Figure A.1(b)).

These observed proportions of clothing adaptation should however be interpreted with caution, since it is possible that small adjustments of clothing level (such as shortening



(a) Observed and fitted clothing levels versus daily mean outdoor temperature

(b) Observed and fitted clothing levels versus indoor temperature

Figure A.2: Clothing level versus indoor and outdoor daily mean temperature

the sleeves or opening the collar of their shirts) are also performed by occupants, whether consciously or not. There may thus exist small but nevertheless significant variations of clothing level of the order of 0.1 clo which may well occur more frequently than the relatively large adjustments that we have observed.

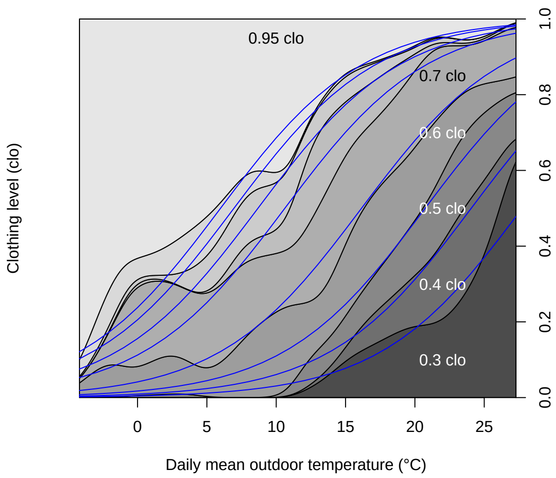
Predicting daily clothing level

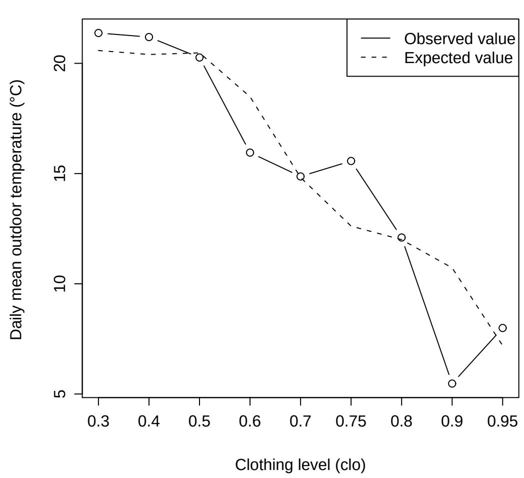
Linear regression. To facilitate comparisons with previous research, we perform linear regression between observed clothing levels and thermal stimuli. Among outdoor temperature and its derivatives, daily mean outdoor temperature produces the best model, far better than indoor temperature:

$$I_{\mathrm{cl}} = 0.929 - 0.0184 \cdot \theta_{\mathrm{out,dm}} \quad (R^2 = 0.405) \tag{A.1}$$
$$I_{\mathrm{cl}} = 1.719 - 0.0445 \cdot \theta_{\mathrm{in}} \quad (R^2 = 0.137) \tag{A.2}$$

Furthermore a model including together $\theta_{\mathrm{out,dm}}$ and θ_{in} does not yield any significant improvement based on analysis of variance ($F = 6.05$, $p = 0.014$) and the fact that the R^2 remains almost unchanged at 0.406. Despite the fact that mean outdoor temperature is a more effective predictor it is evident from Figure A.2 that this model is associated with considerable dispersion which suggests the need for a more informative model formulation.

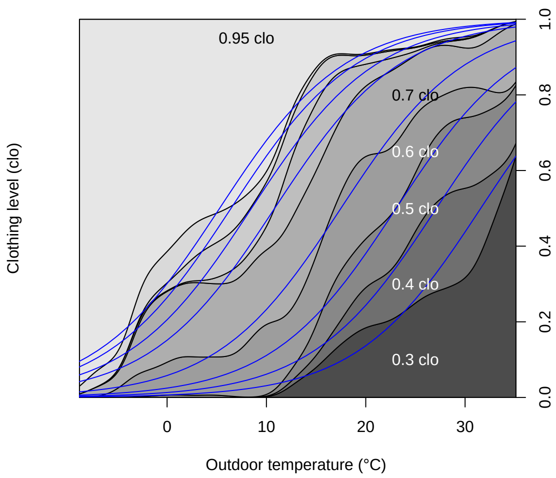
Ordinal logistic regression. Although clothing level is a continuous variable, our questionnaire only allowed us to collect approximate values corresponding to the proposed list of items displayed in Table 2.2. Therefore, we have no choice but to study clothing level as a categorical variable whose ordered discrete levels may be seen as defined by cutpoints applied on an underlying unknown continuous variable. An ordinal logistic model (see Section 3.2.6) is the appropriate approach in this situation.

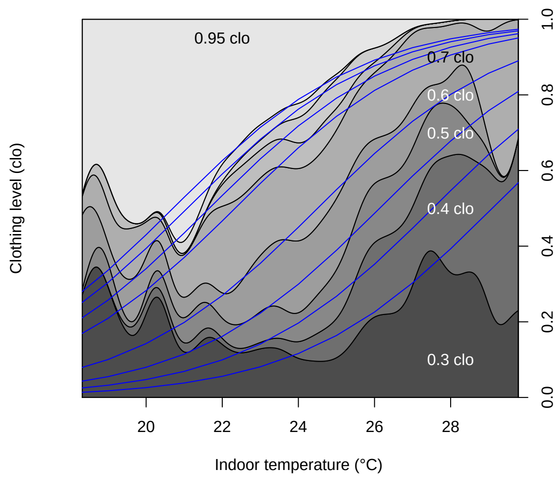




(a) Observed and fitted distributions of clothing levels with respect to daily mean outdoor temperature

(b) Verification of the proportional odds assumption with daily mean outdoor temperature as variable





(c) Observed and fitted distributions of clothing levels with respect to outdoor temperature

(d) Observed and fitted distributions of clothing levels with respect to indoor temperature

Figure A.3: Ordinal logistic regression on clothing level

	LR	**AUC**	$\mathbf{D_{xy}}$	$\mathbf{\Gamma}$	$\tau_{\mathbf{a}}$	$\mathbf{R^2_N}$	**B**
$\theta_{\text{out,dm}}$	1763.9	0.760	0.520	0.523	0.444	0.400	0.104
$\theta_{\text{out,rm}}$	1687.4	0.752	0.504	0.508	0.430	0.387	0.104
θ_{out}	1677.8	0.748	0.497	0.499	0.424	0.386	0.108
$\theta_{\text{out,mm}}$	1485.1	0.731	0.461	0.466	0.394	0.355	0.105
θ_{in}	540.0	0.649	0.298	0.303	0.255	0.145	0.130
$\theta_{\text{out,dm}}$, θ_{in}	1767.5	0.757	0.514	0.515	0.439	0.402	0.112

Table A.1: Goodness-of-fit estimators for logistic models including one or several variables

From Equation (3.17), we express the probability for the clothing level I_{cl} to be superior to a given threshold value I_j as a logistic distribution:

$$p(I_{\text{cl}} \geq I_j) = \frac{\exp(a_j + b \cdot \theta)}{1 + \exp(a_j + b \cdot \theta)}, \tag{A.3}$$

where θ is the selected most relevant available thermal variable, such as θ_{in}, θ_{out} or some combination of them.

We fit ordinal logistic models for each of these predictors (Table A.1), and observe that $\theta_{\text{out,dm}}$ is once again the variable with strongest association (Figure A.3(a)), closely followed by θ_{out} (Figure A.3(c)) and that the model with θ_{in} performs relatively poorly (Figure A.3(d)). In general, averages on θ_{out} offer good performance, except the monthly mean $\theta_{\text{out,mm}}$.

We have also attempted to add θ_{in} as a second variable, but once again its inclusion is not significant according to the likelihood ratio test ($G = 3.52$, $p = 0.06$); it also offers no clear goodness-of-fit improvement. Furthermore other variables perform no better. We keep thus as a final model for clothing choice Equation (A.3) with $\theta = \theta_{\text{out,dm}}$, where $\mathbf{I} = (0.4, 0.5, \ldots, 0.9, 0.95)$, $\mathbf{a} = (5.40 \pm 0.11, 4.68 \pm 0.10, 4.04 \pm 0.10, 3.14 \pm 0.09, 2.07 \pm 0.08, 1.69 \pm 0.08, 1.35 \pm 0.08, 1.17 \pm 0.07)$ and $b = -0.1946 \pm 0.0049$.

Finally we check the ordinality assumption (see Section 3.2.6) by plotting the means of $\theta_{\text{out,dm}}$ versus the levels of I_{cl} (Figure A.3(b)), together with the expected value of $\theta_{\text{out,dm}}$ for each level of I_{cl} (shown as a dotted line) under the proportional odds assumption. The low observed discrepancy confirms the validity of this assumption in this situation.

This model formulation is more informative than ordinary linear regression: it explicitly gives as a result the distribution of predicted clothing levels rather than a mean value and correctly treats the surveyed discrete clothing levels; this being encompassed in a single mathematical expression.

Predicting clothing changes

The very small number of observed transitions in clothing level (57 clothing removals and 5 additions) is problematic for the detection of trends in clothing adaptation. We attempted to fit logistic models to infer action probabilities using a similar approach as in Chapters 4 and 5 and found that there is no significant contribution of any available predictor to explain the variation of the very small observed probabilities.

The rare changes in clothing may be explained by the general satisfaction of the thermal conditions in the building, or by the large availability of other controls to adapt the environment. Another possible explanation may lie in errors in judgement (occupants may have adapted to a cooler than predicted temperature, so that they have erroneously under-clothed themselves).

Further measurements with more precise clothing values would thus be helpful to better understand occupants' adaptation in clothing. We may however hypothesise that determinant predictors should include indoor temperature together with the initial clothing level and its ability to be changed by removing or adding a layer of clothing. This suggests the use of transition probabilities between several clothing states, with $p_{ij} = 0$ for levels $I_{\mathrm{cl},j}$ that cannot be reached with available clothing $I_{\mathrm{cl},i}$.

A.1.3 Discussion

Determinant predictors

Based on our observations, there is evidence that outdoor climate and season in general primarily determines effective clothing level in office environments. In contrast with actions on windows (Chapter 4), there is an intrinsic and general influence of outdoor climate, which directly impacts occupants when clothing decisions are taken. The specific impact of daily mean outdoor temperature is coherent, formulating the influence of the near past in the decision, in conjunction with the fact that changes during the day are rare. The removal or addition of layers of clothing on the other hand have proven to be difficult to predict owing to the lack of supportive data. We may at best offer the seemingly logical conclusion that occupants perform changes (if available) based on local indoor stimuli.

In conclusion, clothing adaptation tends to be more a predictive strategy – the level being set at the beginning of the day, based on prior experience of thermal (especially outdoor) conditions, with opportunities for adaptation during the day being rarely exercised.

Dress codes and specificities

In certain office environments, particularly those in which formal attire is favoured, occupants may also be constrained from adjusting their clothing level. Occupants of our relatively informal case study buildings, however, were able to adjust their clothing level in a rather unconstrained way.

A strict dress code implies a suppression of adaptive opportunity (implying a narrower range of acceptable temperatures). This can be modelled using a modified version of Equation (A.3) where low clothing levels are removed, and thus merged with the lowest available level. In residential environments on the other hand adaptation may be relatively unconstrained, allowing for lower clothing levels if necessary, requiring additional levels in Equation (A.3). Figure A.4 summarises the possible formulation for these environment-specific distributions.

Proposal for a model

Based on these observations, we propose a model formulated as follows:

- The range of possible clothing levels is determined by the type of environment to simulate.

- At the start of each day, based on the mean outdoor temperature of the preceding 24 hours, a clothing level is randomly drawn from the distribution of Equation (A.3).

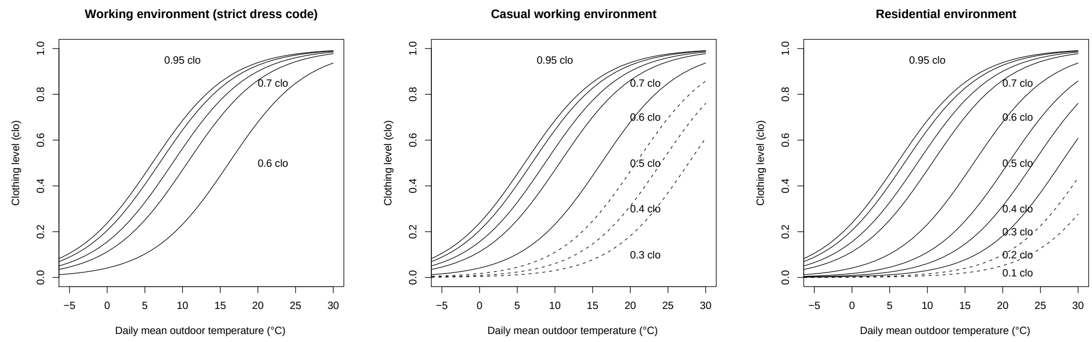


Figure A.4: Hypothetical distributions for clothing in different environments

- For each time step where the occupant is present, additions and removals of clothing are predicted using transition probabilities whose explicit form is to be defined. However, in the absence of data to support this third step the relative rarety of this observed form of adaptation would support an intermediate model based on only the first two steps, which models day-to-day transitions, the most important adaptation mechanism.

A.2 Metabolic activity

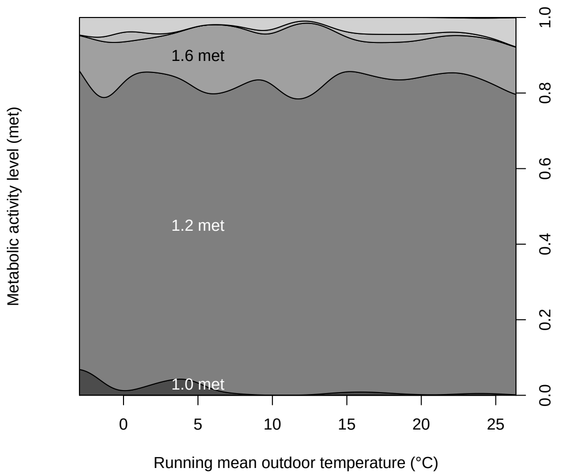
Like clothing, metabolic activity is a determinant variable for thermal comfort and therefore of interest in our research. As with clothing metabolic activity was self-reported by respondents of our electronic questionnaire, in which a choice from a limited set of metabolic activities (seating, standing, etc) was made, see Table 2.2.

A.2.1 State of the art

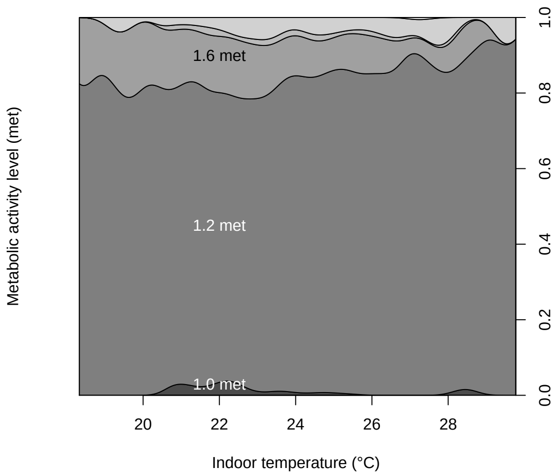
Research on variations of metabolic activity observed on building occupants is very lacunar. A precise measure of this value is also difficult, so that most thermal comfort surveys provide an approximate self-reported value. The question is however of interest, as Humphreys and Nicol [175] have suggested that adaptations in activity rate is one of the numerous behavioural adaptive actions that building occupants can take to restore their comfort in non-optimal conditions.

In his field survey, Rowe [174] observed twice a day six levels of office activities, using weighting factors for the ongoing duration of this activity and for the consumption of snacks, meals, beverages or cigarettes. Based on collected values ranging between 1.0 and 1.9 met, from 20°C to 27°C, it was observed that 78% of respondents reported different activities between morning and afternoon observations. A weak relationship was found between activity rate and indoor operative temperature, but no relationship was found with outdoor temperature.

Using data from a field survey in Tunisia, Bouden and Ghrab [176] found that metabolic rate remained nearly constant between 1.2 and 1.3 met, independent of temperature. They were not able to survey occupants after 2pm, when they leave for their afternoon nap.



(a) Observed activity levels versus daily mean outdoor temperature



(b) Observed activity levels versus indoor temperature

Figure A.5: Observed distribution of activity levels

A.2.2 Results and discussion

As with clothing we attempted to fit an ordinal logistic model for reported categories of metabolic activity levels. But none of the measured variables could explain the very tiny observed variations in activity levels (82.7% of answers indicated "Sedentary activity", corresponding to $M = 1.2$ met). To illustrate the problem, Figure A.5 shows the distribution of activity with respect to $\theta_{\mathrm{out,rm}}$ and θ_{in}.

Occupants may, in principle, adapt activity levels in response to environmental stimuli, likewise their clothing. In contrast with residences, in workplaces our activities tend to be dictated by the tasks in hand. Metabolic activity in offices may be particularly constrained, being essentially sedentary (desk-based) in nature. It is, we suggest, due to this constrained potential that we observe no discernible statistical correlation between the adaptation of activity and any kind of stimuli.

It is however possible that, under extreme thermal conditions, occupants vary the intensity of their desk-based activity, but our electronic questionnaire was not designed to address this question.

Outside of the working environment, occupants have in principle the freedom to adapt their activities according to their personal preferences (which may or may not be decided in response to environmental stimuli). An obvious and well-known example is the siesta which is common to several Mediterranean countries, where activities are slowed down during the peak temperatures of the day. However, this kind of adaptation can probably be considered as a last solution in the range of adaptations to restore thermal comfort. It seems then that there is only marginal interest in modelling variations in metabolic activity as an adaptation response to restore thermal comfort.

A further confounding factor is that the exact metabolic rate is also a function of other adaptive actions, like consuming hot or cold drinks or a meal. This we model separately in the next section.

A.2.3 Use of drinks

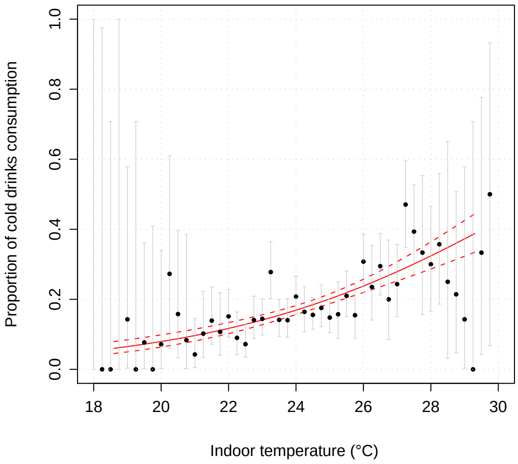
Taking a cold drink was reported for 18.9% of hours preceding the questionnaire (13.5% for hot drinks); it is then a relatively prevalent activity. We have performed logistic regressions for the probability of these events and found that θ_{in} is the best predictor for the consumption of cold drinks, while this is $\theta_{\text{out,rm}}$ in the case of hot drinks. We obtain then:

$$p_{\text{hd}}(\theta_{\text{out,rm}}) = \frac{\exp(a_{\text{hd}} + b_{\text{hd}}\theta_{\text{out,rm}})}{1 + \exp(a_{\text{hd}} + b_{\text{hd}}\theta_{\text{out,rm}})}, \quad a_{\text{hd}} = -6.7 \pm 0.6, \quad b_{\text{hd}} = 0.21 \pm 0.02,$$

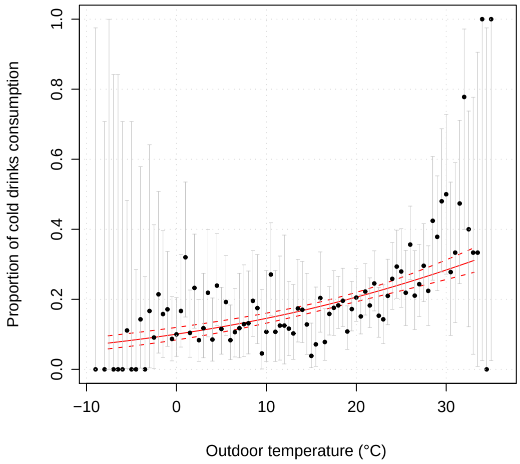
$$p_{\text{cd}}(\theta_{\text{in}}) = \frac{\exp(a_{\text{cd}} + b_{\text{cd}}\theta_{\text{in}})}{1 + \exp(a_{\text{cd}} + b_{\text{cd}}\theta_{\text{in}})}, \quad a_{\text{cd}} = -1.32 \pm 0.09, \quad b_{\text{cd}} = -0.037 \pm 0.006.$$

Cold drink consumption is also reliably modelled by outdoor temperature, perhaps suggesting the presence of some form of seasonal adaptation, rather than being influenced solely by internal conditions. The observed and fitted probabilities are shown in Figure A.6. From this it seems that building occupants consume cold drinks with a greater frequency, relative to a basic minimum (temperature invariant) consumption, when the indoor thermal conditions are warmer; hot drinks are more clearly related with a variable expressing seasonal variations, implying no significant variations during the day.

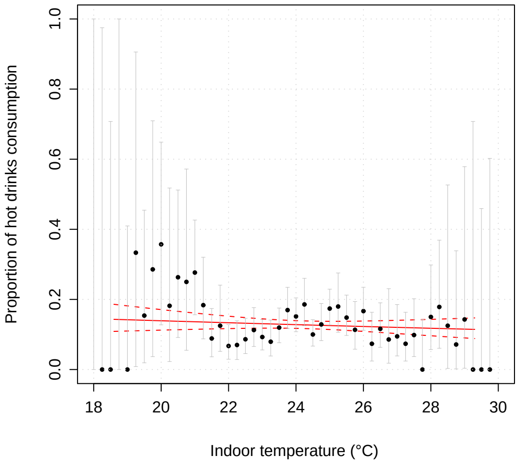
In conclusion, office occupants seem to adapt their metabolic rate in response to thermal stimuli, but this adaptation takes place through the use of drinks; their sole unconstrained means of adaptation in a working environment. However we do not attempt to quantify the impact of the use of drinks on occupants' metabolic activity, as their effect on thermal regulation is complex.



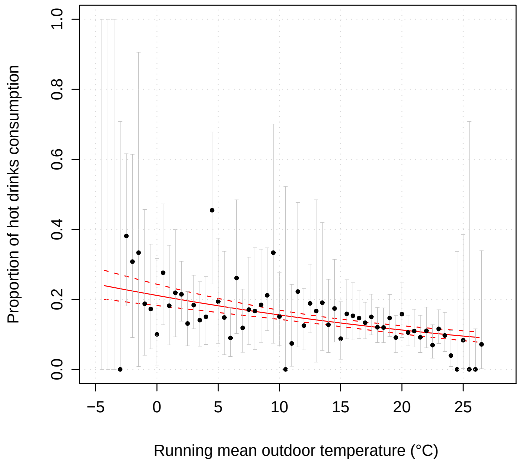
(a) Cold drinks consumption with respect to indoor temperature



(b) Cold drinks consumption with respect to outdoor temperature



(c) Hot drinks consumption with respect to indoor temperature



(d) Hot drinks consumption with respect to exponentially weighted running mean outdoor temperature

Figure A.6: Probability for hot and cold drinks to be consumed during the previous hour

Appendix B

Comfort and its interactions

In this chapter we check for the existence of mutual interactions between thermal, visual and olfactory sensations; more precisely whether the states of comfort and discomfort in one of these aspects has implications for other aspects of comfort probability (as derived in Sections 6.3 and 6.5). From these analyses we show that thermal and visual comfort probabilities generally reach slightly lower values when their counterparts are rated uncomfortable, while olfactory sensation does not display any noticeable effect.

B.1 Previous research

The interaction between thermal, visual and olfactory comfort has thus far been little explored.

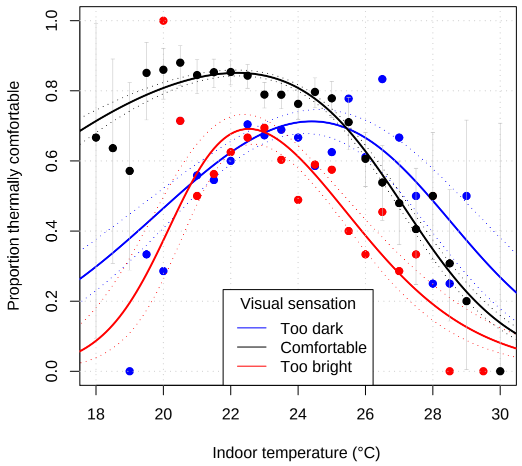
Leaman and Bordass [117] hypothesise that interactions between these different aspects of environmental comfort do exist. Labelled as “revenge effects” they cite the frustration of occupants with interacting problems in buildings, which negatively impacts their assessment of overall indoor environment quality. More specifically that under-performance in one aspect of environmental comfort may negatively impact on others.

Among the rare examples of the use of experimental data to investigate this issue, Roulet et al. [3, 4] observed either global acceptance or global rejection of indoor environment quality based on thermal and visual satisfaction, with an impact on perceived wellbeing. Unfortunately no concomitant measurements of key environmental variables was rigorously performed to support these conclusions.

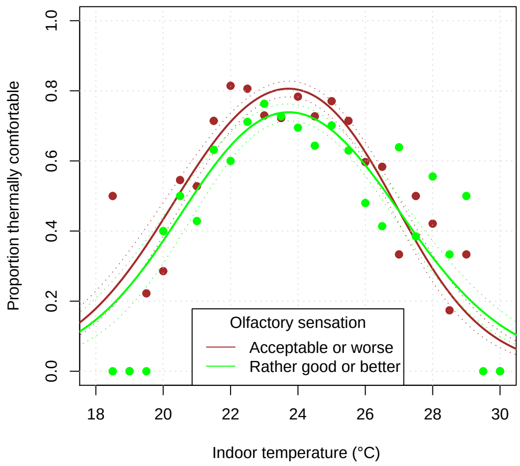
On the other hand, using a designed experiment where twenty subjects evaluated their thermal and visual comfort at two different temperatures (20.5°C and 27°C) for three light source types at a constant 300 lux illuminance, Laurentin et al. [140] failed to observe any influence of thermal conditions on visual comfort appraisal.

B.2 Observations

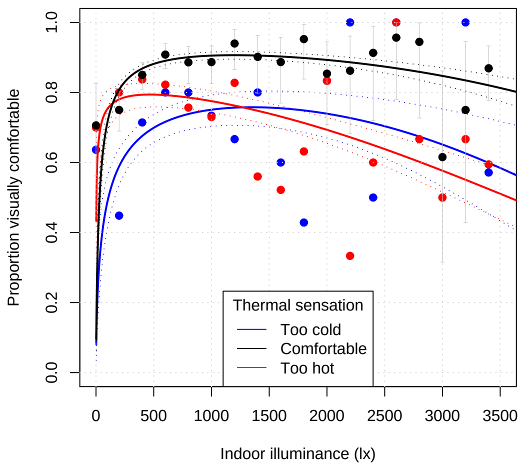
In the first instance we observe that the correlation between θ_{in} and E_{in} (the driving variables for thermal and visual comfort) is low both in the overall dataset ($\rho = 0.150$, Figure 2.6(b)) and in the questionnaire dataset ($\rho = 0.043$). There is thus no problematic correlation between these driving variables that would undermine our study of the impacts of environmental comfort in one domain over that of another.



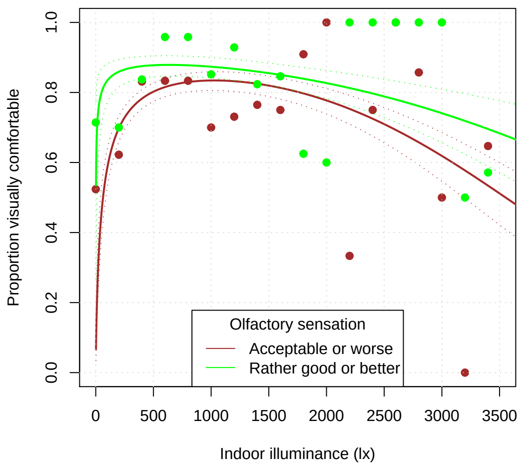
(a) Influence of visual sensation on thermal comfort probability



(b) Influence of olfactive sensation on thermal comfort probability



(c) Influence of thermal sensation on visual comfort probability



(d) Influence of olfactive sensation on visual comfort probability

Figure B.1: Interactions between thermal, visual and olfactory sensations: fitted comfort probabilities with standard errors (dashed lines) and observed proportions of comfortable votes

Impact on thermal comfort. Using the same method as in Section 6.3.2, we derive the probabilities of thermal comfort with respect to visual ($P(S_{\text{th}} = 0|\theta_{\text{in}}, S_{\text{vis}})$) and olfactory sensation ($P(S_{\text{th}} = 0|\theta_{\text{in}}, S_{\text{olf}})$). The obtained distributions are displayed in Figures B.1(a)-B.1(b).

The fitted curves $P(S_{\text{th}} = 0|\theta_{\text{in}}, S_{\text{vis}})$ differ slightly between different visual comfort conditions. It can be noticed that thermal comfort probability decreases more sharply at low temperatures when occupants are visually uncomfortable, and that the maximum reached by $P(S_{\text{th}} = 0|\theta_{\text{in}}, S_{\text{vis}})$ is lower when occupants are visually uncomfortable. In other words, when we are visually comfortable, we are also more thermally comfortable, but when it is too dark we prefer a higher temperature. On the other hand, there is no sign that olfactory sensation influences thermal comfort.

To test the hypothesis of a significant impact of S_{vis} on thermal comfort, we attempted to fit models for p_{cold} and p_{hot} that include S_{vis} as a multiple-level factor, but this latter is not significant according to the Wald test, likewise S_{olf}.

Impact on visual comfort. Similarly, we infer visual comfort probabilities with respect to thermal ($P(S_{\text{vis}} = 0|E_{\text{in}}, S_{\text{th}})$) and olfactory sensation ($P(S_{\text{vis}} = 0|E_{\text{in}}, S_{\text{olf}})$), shown in Figures B.1(c)-B.1(d). We again notice a small deviation in the case of uncomfortable visual conditions with lower maximum comfort probability, so that visual comfort would improve when occupants are thermally comfortable, while visual comfort probability remains practically unchanged in the case of olfactory sensation. A formal fit of models for p_{dark} and p_{bright} including S_{th} and S_{olf} leads to the same conclusions as above.

Olfactory comfort. We do not display any probability of olfactory comfort and do not study the impact of other comfort variables on this latter, as no associated environmental variable could be measured during our survey. However, an approach similar as above could be used in this case if some simultaneous measure of indoor air quality (eg. CO_2 concentration) was available.

B.3 Discussion

Based on the analysis of our data, no clear evidence of interaction between thermal, visual and olfactory sensation could be identified. The only noticeable trend lies in the observation that both visual and thermal comfort probabilities reach a smaller maximum when their counterpart is estimated uncomfortable, although our data cannot provide strong statistical evidence to rigorously support this observation. Nevertheless, further data – particularly from less comfortable buildings – could confirm such a trend, which may be considered as an expression of the "revenge effect" observed by Leaman and Bordass [117].

Another interesting approach would be to also survey global indoor environment acceptance as performed by Wong et al. [177] and precisely assess the impact of each sensation variable and ultimately examine the statistical significance of their interactions on the global outcome.

Appendix C

Productivity of occupants

We made that cut by applying the rule that everything and everybody must produce or get out.
Henry Ford (1863-1947), My Life and Work (1922) [178]

With the decline of agriculture and industry coupled with a rise in the services sector in developed economies, a large part of the population performs their work indoors – a trend reinforced by the development of computer-based work. The link between the indoor environment and productivity at work is thus of increasing economical importance. We investigate such relationships in this final annex.

In this we begin with a short review of previously published research on the link between the indoor environment and productivity at work (Section C.1). Based on our measurements and questionnaire, we present the associations found between perceived productivity and surveyed environmental variables (Section C.2) and discuss the scope and the limitations of these results (Section C.3).

C.1 Introduction

Motivation

Ensuring the conditions leading to highest productivity at work has longsince attracted attention, not only in office environments. The ideas of Henry Ford in the context of industrial environments [178] are amongst the most famous historical examples.

In the office environment, a relationship between indoor environment quality and productivity has longsince been hypothesised. As pointed out by Kosonen and Tan [179, 180], this is an issue of increasing interest since "the salaries of office workers are many times greater than the cost of operating a building in developed countries" and that "there is a potential monetary gain due to improved workers' productivity".

Previous research

Leaman and Bordass [117] performed a questionnaire similar[1] to ours and observed the positive impact on perceived productivity induced by a good degree of perceived personal

[1] The formulation was slightly different: "Please estimate how you productivity at work is increased or decreased by the environmental conditions in the building".

control (on heating, cooling, ventilation and noise) and a rapid environmental response following from control actions.

Several studies concluded that excess temperatures decrease perceived [3, 4] and observed productivity in the context of office [181, 182, 183] and manual work [184]. Based on these observations, Seppanen and Fisk [6] proposed a simple model formulated as a linear productivity decrease of 2% per degree above 25°C. There is less research available on the effects of low temperatures; however Meese et al. [185] found that these cause a reduction of the performance of manual tasks. Jensen et al. [186] also observed that providing good thermal comfort in office environments increases productivity.

Nicol et al. [187] surveyed perceived productivity in offices, in conjunction with a relevant set of physical variables and thermal and visual sensation. In general, observed associations between physical variables were not found to be significant, except with illuminance (negative association). Some weak correlation between productivity and the use of controls was found to be statistically significant: the use of artificial light would have a negative effect while open blinds would have a positive impact. On the other hand, the findings of the field survey conducted in a factory by Juslén et al. [188] found the opposite result that productivity increases with horizontal illuminance on the workplane.

The application of these findings in building design is straightforward with potentially important implications. For instance, in their theoretical study Kosonen and Tan [179, 180] use the previously observed link between environmental comfort and productivity to assess its improvement for different scenarios of building operation (eg. ventilation rate), based on Fanger's predicted percentage of dissatisfied with respect to thermal [106] and olfactory [73] environment.

C.2 Results

Method

In order to detect the set of predictors impacting on productivity, we observe the variation of this latter with respect to all measured variables. This includes all surveyed environmental variables (Table 2.1) but it is also of interest to examine the impact of daily aggregated values of these latter, as occupants were asked to estimate their productivity at the scale of the present day. For this purpose we will also check for the influence of mean, minimal and maximal indoor temperatures during the day and the degree-hours over a threshold value of 25°C during the occupied period. Occupancy-related patterns are potentially of interest, such as first arrival time, total presence duration, day of the week and month. We also study the impact of mean and extreme reported thermal and visual sensation votes, likewise occupants' clothing level.

To do this, we fit a linear model of self-reported productivity to all considered variables, and analyse the association between them using the obtained R^2 and the significance of this effect with the F test. The assessment of the scale of the effect and necessary graphical checks complete the procedure.

Influential variables

Table C.1 shows values of R^2, F and the associated p-values for the fitted linear models using the 545 entries for productivity in our database. We observe significant variations amongst months, and that several environmental variables are significantly associated with

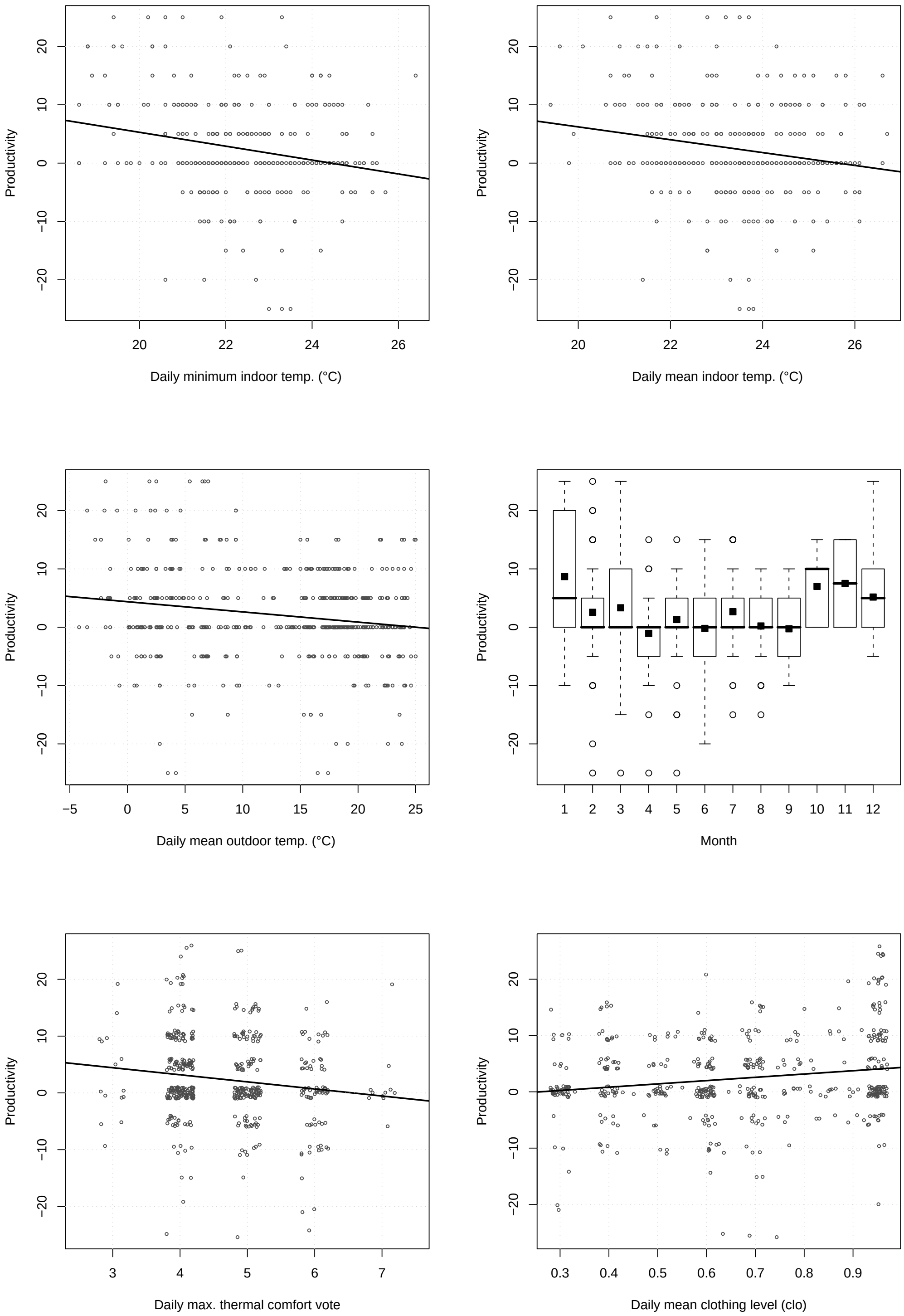


Figure C.1: Perceived productivity versus a selection of significant variables (the points of the bottom charts have been jittered to break ties)

Variable	$\mathbf{R^2}$	F	p-value	Variable	$\mathbf{R^2}$	F	p-value
θ_{out}	0.024	9.241	0.003	$S_{\text{vis,mean}}$	0.003	1.507	0.220
$\theta_{\text{out,dm}}$	0.033	17.871	0.000	$S_{\text{vis,min}}$	0.002	0.980	0.323
$\theta_{\text{out,rm}}$	0.035	19.211	0.000	$S_{\text{vis,max}}$	0.000	0.042	0.838
$\theta_{\text{out,wm}}$	0.036	19.712	0.000	$S_{\text{th,mean}}$	0.005	2.292	0.131
$\theta_{\text{out,mm}}$	0.027	14.536	0.000	$S_{\text{th,min}}$	0.000	0.038	0.845
θ_{in}	0.020	7.766	0.006	$S_{\text{th,max}}$	0.022	11.022	0.001
$\theta_{\text{in,mean}}$	0.036	11.884	0.001	$I_{\text{cl,min}}$	0.036	18.119	0.000
$\theta_{\text{in,max}}$	0.016	5.283	0.022	$I_{\text{cl,max}}$	0.024	12.079	0.001
$\theta_{\text{in,min}}$	0.048	15.789	0.000	$I_{\text{cl,mean}}$	0.032	16.393	0.000
$DH > 25°\text{C}$	0.029	2.371	0.128	T_{pres}	0.005	1.523	0.218
$E_{\text{in,mean}}$	0.000	0.032	0.858	Weekday	0.004	0.467	0.760
First arr.	0.000	0.142	0.706	Month	0.078	3.940	0.000

Table C.1: Values for R^2, F and p-value based on F test obtained from linear models for productivity with each listed variable

a decrease in perceived productivity (θ_{out} and its averages, θ_{in} and its mean, minimum and maximum and the daily maximum thermal sensation $S_{\text{th,max}}$) while $I_{\text{cl,mean}}$, $I_{\text{cl,min}}$ and $I_{\text{cl,max}}$ grow together with productivity; which suggests that we are more productive during colder outside conditions as I_{cl} and θ_{out} are inversely correlated (Section A.1). These trends are visible in Figure C.1.

Variables linked with indoor temperature ($\theta_{\text{in,mean}}$ and $\theta_{\text{in,min}}$) produce the highest values of R^2. These results indicate that the thermal history of the current day has the heaviest impact on productivity: a daily mean indoor temperature of 20°C would increase productivity by 6% compared to 26°C. The use of the daily minimum indoor temperature $\theta_{\text{in,min}}$ results in a slightly higher quality of adjustment than $\theta_{\text{in,mean}}$.

The addition of any individually significant variable to $\theta_{\text{in,min}}$ or $\theta_{\text{in,mean}}$ in a model with two parameters is not statistically significant. We can conclude that the observed association between productivity and variables such as θ_{out}, $I_{\text{cl,mean}}$ and $S_{\text{th,max}}$ is due to their own correlation with indoor thermal stimuli and that they do not display an intrinsic effect. We did not find any significant association with visual stimuli and sensation.

C.3 Discussion

We have isolated the clear adverse influence of a hot indoor environment on occupants' perceived productivity as observed in previous studies. However, the surveyed indoor conditions do not include uncomfortably cold indoor conditions, where a decrease in productivity may be expected. It can thus be hypothesised that curvilinear relations, including for instance polynomial terms (although these appear not to be significant according to our data), would better describe the association between productivity and temperature or daily extreme comfort votes; but this would need confirmation from further observations.

We point out that surveyed variations of perceived productivity might not fully reflect increases or decreases to actual performance. However, assuming coherent answers from the surveyed subjects, we may assume that these latter are proportional.

A large number of other factors are likely to impact productivity, such as access to environmental controls, but our survey did not address these issues. Non-environmental factors, such as management, motivation, office size and occupancy density are possibly more important, so that a large database including a wide range of building types would be necessary to understand all the key mechanisms influencing office productivity. We also underline that the obtained effect is valid in the context of usual office tasks and it is not necessarily valid for other work activities.

Finally, we point out that we have merely found evidence for a significant association, but not attempted to provide a predictive model or any form of validation; the spread being too large and the explained variance too small for such purposes.

Bibliography

[1] European Commission. *European Union Energy and Transport in Figures.* Directorate-General Energy and Transport, Brussels, 2008.

[2] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.

[3] Claude-Alain Roulet, Niklaus Johner, Flavio Foradini, Philomena Bluyssen, Chrit Cox, Eduardo de Oliveira Fernandes, Birgit Müller, and Claire Aizlewood. Perceived health and comfort in relation to energy use and building characteristics. *Building Research and Information*, 34(5):467–474, 2006.

[4] Claude-Alain Roulet, Flourentzos Flourentzou, Flavio Foradini, Philomena Bluyssen, Chrit Cox, and Claire Aizlewood. Multicritera analysis of health, comfort and energy efficiency in buildings. *Building Research and Information*, 34(5):475–482, 2006.

[5] William J. Fisk. How IEQ affects health, productivity. *ASHRAE Journal*, 44(5):56–60, 2002.

[6] O. A. Seppänen and William J. Fisk. Some quantitative relations between indoor environmental quality and work performance or health. *HVAC and R Research*, 12(4):957–973, 2006.

[7] J. A. Clarke. *Environmental systems performance.* PhD thesis, 1977.

[8] M. C. B. Gough. *Modelling heat flow in buildings: An eigenfunction approach.* PhD thesis, 1982.

[9] D. Tang. *Modelling of heating and air-conditioning systems.* PhD thesis, 1984.

[10] J. L. M. Hensen. *On the thermal interaction with building structure and heating and ventilating systems.* PhD thesis, 1991.

[11] B. H. Bland. Conduction in dynamic thermal models: Analytical tests for validation. *Building Services Engineering Research and Technology*, 3(4):197–208, 1992.

[12] R. Judkoff and J.A. Neymark. A testing and diagnostic procedure for building energy simulation programs. In *BEP '94*, York, United Kingdom, 1995.

[13] K. J. Lomas, H. Eppel, C. J. Martin, and D. P. Bloomfield. Empirical validation of building energy simulation programs. *Energy and Buildings*, 26(3):253–267, 1997.

[14] A. Nakhi. *Adaptive construction modelling within whole building dynamic simulation.* PhD thesis, 1995.

[15] M. Janak. Coupling building and lighting simulation. In *Building Simulation 1997: Fifth International IBPSA Conference*, Prague, Czech Republic, 1997.

[16] N. Kelly. *Towards A Design Environment For Building Integrated Energy Systems: The Integration Of Electrical Power Flow Modelling With Building Simulation*. PhD thesis, 1998.

[17] Ian Beausoleil-Morrison. *The adaptive coupling of heat and airflow modelling within dynamic whole-building simulation*. PhD thesis, 2000.

[18] Kevin J. Lomas and Herbert Eppel. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, 19:21–44, 1992.

[19] Iain Macdonald and Paul Strachan. Practical application of uncertainty analysis. *Energy and Buildings*, 33(3):219–227, 2001.

[20] J. A. Clarke. *Energy Simulation in Building Design, Second Edition*. Butterworth-Heinemann, Oxford, United Kingdom, 2001.

[21] Nick Baker. Low energy strategies. In *Energy and environment in non-domestic buildings*. Cambridge Architectural Research Ltd, Cambridge, United Kingdom, 1994.

[22] Clive Seligman, John M. Darley, and Lawrence J. Becker. Behavioral approaches to residential energy conservation. *Energy and Buildings*, 1(3):325–337, 1977/78.

[23] C. Dubrul. Inhabitant behaviour with respect to ventilation – A summary report of IEA Annex VIII. Technical report, 1988.

[24] Go Iwashita and Hiroshi Akasaka. The effects of human behavior on natural ventilation rate and indoor air environment in summer – a field study in southern Japan. *Energy and Buildings*, 25(3):195–205, 1997.

[25] A. S. Bahaj and P. A. B. James. Urban energy generation: The added value of photovoltaics in social housing. *Renewable and Sustainable Energy Reviews*, 11(9):2121–2136, 2007.

[26] P. Hoes, J. L. M. Hensen, M. G. L. C. Loomans, B. de Vries, and D. Bourgeois. User behaviour in whole building simulation. *Energy and Buildings*, 2009.

[27] Danni Wang, Clifford C. Federspiel, and Francis Rubinstein. Modelling occupancy in single person offices. *Energy and Buildings*, 37(2):121–126, 2005.

[28] Jessen Page. *Simulating occupant presence and behaviour in buildings*. PhD thesis, 2007.

[29] Jessen Page, Darren Robinson, Nicolas Morel, and Jean-Louis Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40(2):83–98, 2008.

[30] Ian Richardson, Murray Thomson, and David Infield. A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40(8):1560–1566, 2008.

[31] John W. Tukey. Sunset Salvo. *The American Statistician*, 40(1):72–76, 1986.

[32] René Altherr and Jean-Bernard Gay. A low environmental impact anidolic façade. *Building and Environment*, 37(12):1409–1419, 2002.

[33] David Lindelöf. *Bayesian optimization of visual comfort*. PhD thesis, 2007.

[34] P. Kenneth Seidelmann. *Explanatory Supplement to the Astronomical Almanac*. University Science Books, Mill Valley, California, 1992.

[35] International Organisation for Standardisation. Moderate thermal environments – Determination of the PMV and PPD indices and specification of the conditions for thermal comfort, 1994.

[36] Annette J. Dobson. *An Introduction to Generalized Linear Models, Second Edition*. Chapman & Hall, 2002.

[37] David W. Hosmer and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, New York, USA, 2nd edition, 2000.

[38] Anthony C. Davison. *Statistical Models*. Cambridge University Press, Cambridge, United Kingdom, 2003.

[39] W. W. Hauck and A. Donner. Wald's test as applied to hypotheses in logit analyses. *Journal of the American Statistical Association*, 72:851–853, 1977.

[40] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[41] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[42] David W. Hosmer, T. Hosmer, S. Le Cessie, and Stanley Lemeshow. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16:965–980, 1997.

[43] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):882–891, 2006.

[44] N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.

[45] Frank E. Harrel. *Regression Modeling Strategies*. Springer Series in Statistics. Springer, 2001.

[46] Morten W. Fagerland, David W. Hosmer, and Anna M. Bofin. Mutlinomial goodness-of-fit tests for logistic regression models. *Statistics in Medicine*, 27:4238–4253, 2008.

[47] Alan Agresti. *Categorical Data Analysis*. Wiley, New York, 1990.

[48] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, United Kingdom, 1997.

[49] P. J. Avery and D. A. Henderson. Fitting Markov chain models to discrete state series such as DNA sequences. *Applied Statistics*, 48:53–61, 1999.

[50] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[51] David G. Kleinbaum and Mitchel Klein. Survival Analysis – A Self-Learning Text, 2005.

[52] Frédéric Haldi and Darren Robinson. A comprehensive stochastic model of window usage: Theory and validation. In *Building Simulation 2009: 11th International IBPSA Conference*, pages 545–552, Glasgow, United Kingdom, 2009.

[53] Frédéric Haldi and Darren Robinson. Interactions with window openings by office occupants. *Building and Environment*, 44(12):2378–2395, 2009.

[54] J. B. Dick and D. A. Thomas. Ventilation research in occupied houses. *Journal of the Institution of Heating and Ventilating Engineers*, 19:306–326, 1951.

[55] G. W. Brundrett. Ventilation: A behavioural approach. *Energy Research*, 1:289–298, 1977.

[56] G. W. Brundrett. Window ventilation and human behaviour. In P. O. Fanger and O. Valbjorn, editors, *Indoor Climate*, pages 317–330, 1979.

[57] M. D. Lyberg. Energy losses due to airing by occupants. In *CIB Commission W67 Symposium – Energy Conservation in the Built Environment*, Dublin, Ireland, 1982.

[58] P. R. Warren and L. M. Parkins. Window-Opening Behavior in Office Buildings. *ASHRAE Transactions*, 90(1B):1056–1076, 1984.

[59] R. Fritsch, A. Kohler, M. Nygård-Ferguson, and J.-L. Scartezzini. A Stochastic Model of User Behaviour Regarding Ventilation. *Building and Environment*, 25(2):173–181, 1990.

[60] C.-A. Roulet, P. Cretton, R. Fritsch, and J.-L. Scartezzini. Stochastic Model of Inhabitant Behavior with Regard to Ventilation. Technical report, 1991.

[61] H. B. Awbi. *Ventilation of Buildings.* E and FN Spon, 1995.

[62] J. Fergus Nicol and S. C. Roaf. Pioneering new indoor temperature standards: the Pakistan Project. *Energy and Buildings*, 23:169–174, 1996.

[63] J. Fergus Nicol, Iftikhar A. Raja, Arif Allaudin, and Gul Najam Jamy. Climatic variations in comfortable temperatures: the Pakistan projects. *Energy and Buildings*, 30(3):261–279, 1999.

[64] Iftikhar A. Raja, J. Fergus Nicol, and Kathryn J. McCartney. Natural ventilated buildings: use of controls for changing indoor climate. *Renewable Energy*, 15:391–394, 1998.

[65] I. A. Raja, J. Fergus Nicol, K. J. McCartney, and M. A. Humphreys. Thermal comfort: use of controls in naturally ventilated buildings. *Energy and Buildings*, 33(3):235–244, 2001.

[66] Kathryn J. McCartney and J. Fergus Nicol. Developing an adaptive control algorithm for Europe. *Energy and Buildings*, 34(6):623–635, 2002.

[67] J. Fergus Nicol. Characterising occupant behaviour in buildings: Towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans. In *Seventh International IBPSA Conference Proceedings*, Rio de Janeiro, 2001.

[68] J. Fergus Nicol and Michael A. Humphreys. A stochastic approach to thermal comfort – Occupant behaviour and energy use in buildings. *ASHRAE Transactions*, 110(2):554–568, 2004.

[69] Darren Robinson. Some trends and research needs in energy and comfort prediction. In *Comfort and energy use in buildings: Getting them right*, Windsor, United Kingdom, 2006.

[70] Hom B. Rijal, Paul Tuohy, Michael A. Humphreys, J. Fergus Nicol, Aizaz Samuel, and J. Clarke. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. *Energy and Buildings*, 39(7):823–836, 2007.

[71] Hom B. Rijal, Paul Tuohy, J. Fergus Nicol, Michael A. Humphreys, Aizaz Samuel, and J. Clarke. Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings. *Journal of Building Performance Simulation*, 1(1):17–30, 2008.

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

[73] P. Ole Fanger. Introduction of the olf and the decipol units to quantify air pollution perceived by humans indoors and outdoors. *Energy and Buildings*, 12(1):1–6, 1988.

[74] Frédéric Haldi and Darren Robinson. On the behaviour and adaptation of office occupants. *Building and Environment*, 43(12):2163–2177, 2008.

[75] Geun Young Yun and Koen Steemers. User behaviour of window control in offices during summer and winter. In *CISBAT Conference*, Lausanne, Switzerland, 2007.

[76] Geun Young Yun and Koen Steemers. Time-dependent occupant behaviour models of window control in summer. *Building and Environment*, 43(9):1471–1482, 2008.

[77] Hom B. Rijal, Paul Tuohy, Michael A. Humphreys, J. Fergus Nicol, Aizaz Samuel, Iftikhar A. Raja, and Joe Clarke. Development of Adaptive Algorithms for the Operation of Windows, Fans and Doors to Predict Thermal Comfort and Energy Use in Pakistani Buildings. *ASHRAE Transactions*, 114(2):555–573, 2008.

[78] Christoph Reinhart and K. Voss. Monitoring manual control of electric lighting and blinds. *Lighting Research and Technology*, 35(3):243–260, 2003.

[79] C. Reinhart. Lighswitch-2002: a model for manual and automated control of electric lighting and blinds. *Solar Energy*, 77:15–28, 2004.

[80] Geun Young Yun, Paul Tuohy, and Koen Steemers. Thermal performance of a naturally ventilated building using a combined algorithm of probabilistic occupant behaviour and deterministic heat and mass balance model. *Energy and Buildings*, 41(5):489–499, 2009.

[81] Sebastian Herkel, Ulla Knapp, and Jens Pfafferott. A preliminary model of user behaviour regarding the manual control of windows in office buildings. In *Building Simulation 2005, Ninth International IBPSA Conference*, Montréal, Canada, 2005.

[82] Sebastian Herkel, Ulla Knapp, and Jens Pfafferott. Towards a model of user behaviour regarding the manual control of windows in office buildings. *Building and Environment*, 43(4):588–600, 2008.

[83] Ardeshir Mahdavi and Claus Pröglhöf. A model-based approach to natural ventilation. *Building and Environment*, 43:620–627, 2008.

[84] Ardeshir Mahdavi, Elham Kabir, Abdolazim Mohammad, and Claus Pröglhöf. User-based window operation in an office building. In *Indoor Air 2008*, Copenhagen, Denmark, 2008.

[85] Rune Vinther Andersen, Jørn Toftum, Klaus Kaae Andersen, and Bjarne W. Olesen. Survey of occupant behaviour and control of indoor environment in Danish dwellings. *Energy and Buildings*, 41(1):11–16, 2009.

[86] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[87] Frank E. Harrell Jr. *Design Package*, 2007. R package version 2.1.1.

[88] Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–117, 2001.

[89] Frédéric Haldi and Darren Robinson. A comprehensive stochastic model of blind usage: Theory and validation. In *Building Simulation 2009: 11th International IBPSA Conference*, pages 529–536, Glasgow, United Kingdom, 2009.

[90] Frédéric Haldi and Darren Robinson. Adaptive actions on shading devices in response to local visual stimuli. *Journal of Building Performance Simulation*, 2010.

[91] M. S. Rea. Window blind occlusion: a pilot study. *Building and Environment*, 19(2):133–137, 1984.

[92] T. Inoue, T. Kawase, T. Ibamoto, S. Takakusa, and Y. Matsuo. The development of an optimal control system for window shading devices based on investigations in office buildings. *ASHRAE Transactions*, 94(2):1034–1049, 1988.

[93] Yannick Sutter, Dominique Dumortier, and Marc Fontoynont. The use of shading systems in VDU task offices: A pilot study. *Energy and Buildings*, 38(7):780–789, 2006.

[94] David Lindelöf and Nicolas Morel. Bayesian estimation of visual discomfort. *Building Research and Information*, 36(1):83–96, 2008.

[95] Ardeshir Mahdavi, Abdolazim Mohammadi, Elham Kabir, and Lyudmila Lambeva. Occupants' operation of lighting and shading systems in office buildings. *Journal of Building Performance Simulation*, 1(1):57–65, 2008.

[96] Ardeshir Mahdavi, Abdolazim Mohammadi, Elham Kabir, and Lyudmila Lambeva. Shading and Lighting Operation in Office Buildings in Austria: A Study of User Control Behavior. *Building Simulation: An International Journal*, 1(2):111–117, 2008.

[97] Ardeshir Mahdavi and Claus Pröglhöf. Observation-based models of user control actions in buildings. In *PLEA 2008 – 25th Conference on Passive and Low Energy Architecture*, Dublin, Ireland, 2008.

[98] Vorapat Inkarojrit. Monitoring and modelling of manually-controlled Venetian blinds in private offices: a pilot study. *Journal of Building Performance Simulation*, 1(2):75–89, 2008.

[99] H. D. Einhorn. A new method for the assessment of discomfort glare. *Lighting Research and Technology*, 1(4):235–247, 1969.

[100] H. D. Einhorn. Discomfort glare: a formula to bridge differences. *Lighting Research and Technology*, 11(2):90–94, 1979.

[101] CIE. Discomfort glare in the interior lighting, 1992.

[102] Jan Wienold and Jens Christoffersen. Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. *Energy and Buildings*, 38(7):743–757, 2006.

[103] Darren Robinson and Frédéric Haldi. Preliminary model of overheating risk based on field survey data. In *10th IBPSA Conference*, pages 745–750, Beijing, China, 2007.

[104] Frédéric Haldi and Darren Robinson. Representing behaviour and adaptation of office occupants in building simulation. In *CISBAT Conference*, pages 319–324, Lausanne, 2007.

[105] ASHRAE Standard 55-2004 – Thermal Environmental Conditions for Human Occupancy, 2004.

[106] P. Ole Fanger. *Thermal comfort: Analysis and Applications in Environmental Engineering*. Danish Technical Press, 1970.

[107] L. G. Berglund. Thermal acceptability. *ASHRAE Transactions*, 85(2):825–834, 1979.

[108] Yufeng Zhang and Rongyi Zhao. Overall thermal sensation, acceptability and comfort. *Building and Environment*, 43(1):44–50, 2008.

[109] Michael A. Humphreys and J. Fergus Nicol. Do people like to feel neutral? Response to the ASHRAE scale of subjective warmth in relation to thermal preference, indoor and outdoor temperature. *ASHRAE Transactions*, 110(2):569–577, 2004.

[110] Arden L. Buck. New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, 20(12):1527–1532, 1981.

[111] Max Sherman. A simplified model of thermal comfort. *Energy and Buildings*, 8(1):37–50, 1985.

[112] Guy R. Newsham. Clothing as a thermal comfort moderator and the effect on energy consumption. *Energy and Buildings*, 26(3):283–291, 1997.

[113] Bjarne W. Olesen and K. C. Parsons. Introduction to thermal comfort standards and to the proposed new version of EN ISO 7730. *Energy and Buildings*, 34:537–548, 2002.

[114] Michael A. Humphreys and J. Fergus Nicol. The validity of ISO-PMV for predicting comfort votes in every-day thermal environments. *Energy and Buildings*, 34(6):667–684, 2002.

[115] J. van Hoof. Forty years of Fanger's model of thermal comfort: comfort for all? *Indoor Air*, 18:182–201, 2008.

[116] Nigel A. Oseland. Predicted and reported thermal sensation in climate chambers, offices and homes. *Energy and Buildings*, 23(2):105–115, 1995.

[117] A. Leaman and B. Bordass. Productivity in buildings: the 'killer' variables. *Building Research and Information*, 27(1):4–20, 1999.

[118] Gail S. Brager, G. Paliaga, and Richard De Dear. Operable windows, personal control, and occupant comfort. *ASHRAE Transactions*, 110(2):17–35, 2004.

[119] Sami Karjalainen. Thermal comfort and use of thermostats in Finnish homes and offices. *Building and Environment*, 2009.

[120] J. Fergus Nicol and Michael A. Humphreys. Thermal comfort as part of a self-regulating system. *Building Research and Practice*, 6(3):174–179, 1973.

[121] J. Fergus Nicol and Michael A. Humphreys. Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251. *Building and Environment*, 2009.

[122] European standard prENrev 15251:2006, Indoor environmental input parameters for design and assessment of energy performance of buildings – addressing indoor air quality, thermal environment, lighting and acoustics, 2006.

[123] Chartered Institution of Building Services Engineers. CIBSE Guide A, 7th edition, 2006.

[124] Richard J. de Dear and Gail S. Brager. Thermal comfort in naturally ventilated buildings : revisions to ASHRAE Standard 55. *Energy and Buildings*, 34(6):549–561, 2002.

[125] Richard J. De Dear and Gail S. Brager. Developing an adaptive model of thermal comfort and preference. *ASHRAE Transactions*, 104:145–167, 1998.

[126] Richard J. De Dear and Gail S. Brager. The adaptive model of thermal comfort and energy conservation in the built environment. *International Journal of Biometeorology*, 45:100–108, 2001.

[127] J. A. J. Stolwijk and J. D. Hardy. Temperature Regulation in Man. *Pflügers Archiv*, 291:129–162, 1966.

[128] A. P. Gagge, J. A. J. Stolwijk, and J. D. Hardy. Comfort and thermal sensations and associated physiological responses at various ambient temperatures. *Environmental Research*, 1(1):1–20, 1967.

[129] Hui Zhang. *Human Thermal Sensation and Comfort in Transient and Non-Uniform Thermal Environments.* PhD thesis, 2003.

[130] Charlie Huizenga, Zhang Hui, and Edward Arens. A model of human physiology and comfort for assessing complex thermal environments. *Building and Environment*, 36(6):691–699, 2001.

[131] Dusan Fiala. *Dynamic Simulation of Human Heat Transfer and Thermal Comfort.* PhD thesis, 1998.

[132] Dusan Fiala, Kevin J. Lomas, and Martin Stohrer. Computer prediction of human thermoregulatory and temperature responses to a wide range of environmental conditions. *International Journal of Biometeorology*, 45(3):143–159, 2001.

[133] Christoph van Treeck, Petra Wenisch, André Borrmann, Michael Pfaffinger, Oliver Wenisch, and Ernst Rank. ComfSim – Interaktive Simulation des thermischen Komforts in Innenräumen auf Höchstleistungsrechnern. *Bauphysik*, 29(1):2–7, 2007.

[134] Paul C. Cropper, Tong Yang, Malcolm J. Cook, Dusan Fiala, and Rehan Yousaf. Exchange of simulation data between CFD programmes and a multisegmented human thermal comfort model. In *Air Conditioning and the Low Carbon Cooling Challenge*, Windsor, United Kingdom, 2008.

[135] Paul C. Cropper, Tong Yang, Malcolm J. Cook, Dusan Fiala, and Rehan Yousaf. Simulating the effect of complex indoor environmental conditions on human thermal comfort. In *Building Simulation 2009: 11th International IBPSA Conference*, pages 1367–1373, Glasgow, United Kingdom, 2009.

[136] J. P. Rugh, R. B. Farrington, D. Bharathan, A. Vlahinos, R. Burke, C. Huizenga, and H. Zhang. Predicting human thermal comfort in a transient nonuniform thermal environment. *European Journal of Applied Physiology*, 92(6):721–727, 2004.

[137] H. Zhang, C. Huizenga, E. Arens, and D. Wang. Thermal sensation and comfort in transient non-uniform thermal environments. *European Journal of Applied Physiology*, 92(6):728–733, 2004.

[138] Anca D. Galasiu and Jennifer A. Veitch. Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review. *Energy and Buildings*, 38(7):728–742, 2006.

[139] B. W. P. Wells. Subjective responses to the lighting installation in a modern office building and their design implications. *Building Science*, 1(1):57–68, 1965.

[140] C. Laurentin, V. Berrutto, and Marc Fontoynont. Effect of thermal conditions and light source type on visual comfort appraisal. *Lighting Research and Technology*, 32(4):223–233, 2000.

[141] A. Nabil and J. Mardaljevic. Useful daylight illuminance: a new paradigm for assessing daylight in buildings. *Lighting Research and Technology*, 37(1):41–59, 2005.

[142] I. Griffiths. Thermal comfort studies in buildings with passive solar features, field studies. UK: Report of the Commission of the European Community ENS35 090. Technical report, 1990.

[143] N. V. Baker and M. Standeven. Thermal comfort for free-running buildings. *Energy and Buildings*, 23(3):175–182, 1996.

[144] N. A. Oseland, M. A. Humphreys, J. F. Nicol, N. V. Baker, and K. C. Parsons. Building design and management for thermal comfort, BRE Client Report CR 203/98. Technical report, 1998.

[145] Sustainable Urban Development Group, Solar Energy and Building Physics Laboratory, EPFL. http://lesowww.epfl.ch/e/research urbdev.html.

[146] D. R. G. Hunt. Lighting controls: their current use and possible improvements. *Energy Research*, 2:343–374, 1978.

[147] D. R. G. Hunt. The use of artificial lighting in relation to daylight levels and occupancy. *Building and Environment*, 14(1):21–33, 1979.

[148] D. R. G. Hunt. Predicting artificial lighting use – A method based upon observed patterns of behaviour. *Lighting research and technology*, 12(1):7–14, 1980.

[149] P. Haves and P. Littlefair. Daylight in dynamic thermal modelling programs: Case study. *Building services engineering research and technology*, 9(4):183–188, 1982.

[150] S. Pigg, M. Eilers, and J. Reed. Behavioural aspects of lighting and occupancy sensors in private offices: a case study of a university office building. In *ACEEE Summer Study on Energy Efficiency in Buildings*, volume 8, pages 161–171, 1996.

[151] Christoph Reinhart. *Daylight Availability and Manual Lighting Control in Office Buildings – Simulation Studies and Analysis of Measurements*. PhD thesis, 2001.

[152] Melody Stokes, Mark Rylatt, and Kevin Lomas. A simple model of domestic lighting demand. *Energy and Buildings*, 36(2):103–116, 2004.

[153] Koen Steemers and Geun Young Yun. Household energy consumption: a study of the role of occupants. *Building Research and Information*, 37(5-6):625–637, 2009.

[154] Jukka V. Paatero and Peter D. Lund. A model for generating household electricity load profiles. *International Journal of Energy Research*, 30:273–290, 2006.

[155] Jun Tanimoto and Aya Hagishima. State transition probability for the Markov model dealing with on/off cooling schedule in dwellings. *Energy and Buildings*, 37(3):181–187, 2005.

[156] Marcel Schweiker and Masanori Shukuya. Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions. *Building and Environment*, 44(10):2137–2149, 2009.

[157] Darren Robinson, Frédéric Haldi, Jérôme Henri Kämpf, Philippe Leroux, Diane Perez, Adil Rasheed, and Urs Wilke. CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning. In *Building Simulation 2009: 11th International IBPSA Conference*, pages 1083–1090, Glasgow, United Kingdom, 2009.

[158] Jérôme Henri Kämpf and Darren Robinson. A simplified thermal model to support analysis of urban resource flows. *Energy and Buildings*, 39(4):445–453, 2007.

[159] Darren Robinson and Andrew Stone. Solar radiation modelling in the urban context. *Solar Energy*, 77(3):295–309, 2004.

[160] Darren Robinson and Andrew Stone. Internal illumination prediction based on a simplified radiosity algorithm. *Solar Energy*, 80(3):260–267, 2006.

[161] Jérôme Henri Kämpf. *On the modelling and optimisation of urban energy fluxes*. PhD thesis, 2009.

[162] MATSim-T, 1990. http://www.matsim.org.

[163] Jun Tanimoto, Aya Hagishima, and Hiroki Sagara. Validation of probabilistic methodology for generating actual inhabitants' behavior schedules for accurate prediction of maximum energy requirements. *Energy and Buildings*, 40(3):316–322, 2008.

[164] Jun Tanimoto, Aya Hagishima, and Hiroki Sagara. A methodology for peak energy requirement considering actual variation of occupants' behavior schedules. *Building and Environment*, 43(4):610–619, 2008.

[165] Vincent Tabak and Bauke de Vries. Methods for the prediction of intermediate activities by office occupants. *Building and Environment*, Article in press, 2010.

[166] Sheldon M. Ross. *Simulation, Fourth Edition*. Academic Press, 2006.

[167] Craig Morgan and Richard de Dear. Weather, clothing and thermal adaptation to indoor climate. *Climate Research*, 24:267–284, 2003.

[168] Bjarne W. Olesen. A new simpler method for estimating the thermal insulation of a clothing ensemble. *ASHRAE Transactions*, 91(2B):478–492, 1985.

[169] Michael A. Humphreys. Clothing and thermal comfort of secondary school children in summertime. In *CIB Commission W45 Symposium*, Thermal comfort and moderate heat stress, 1972.

[170] Michael A. Humphreys. Classroom temperature, clothing and thermal comfort – a study of secondary school children in summertime. *Journal of the Institution of Heating and Ventilating Engineers*, 41:191–202, 1973.

[171] Michael A. Humphreys. Clothing and the Outdoor Microclimate in Summer. *Building and Environment*, 12(3):137–142, 1977.

[172] Michael A. Humphreys. The influence of season and ambient temperature on human clothing behaviour. In P. O. Fanger and O. Valbjorn, editors, *Indoor Climate*, pages 699–713, 1979.

[173] Guy R. Newsham and Dale K. Tiller. A field study of office thermal comfort using questionnaire software. *ASHRAE Transactions*, 103(2):3–17, 1997.

[174] David Malcolm Rowe. Activity rates and thermal comfort of office occupants in Sydney. *Journal of Thermal Biology*, 26(4-5):415–418, 2001.

[175] Michael A. Humphreys and F. Nicol. Understanding the adaptive approach to thermal comfort. *ASHRAE Tech. Data Bull.*, 14(1):1–14, 1998.

[176] C. Bouden and N. Ghrab. An adaptive thermal comfort model for the Tunisian context: a field study results. *Energy and Buildings*, 37(9):952–963, 2005.

[177] L. T. Wong, K. W. Mui, and P. S. Hui. A multivariate-logistic model for acceptance of indoor environmental quality (IEQ) in offices. *Building and Environment*, 43(1):1–6, 2008.

[178] Henry Ford and Samuel Crowther. *My Life and Work*. Garden City Publishing Company, New York, 1922.

[179] R. Kosonen and F. Tan. The effect of perceived indoor air quality on productivity loss. *Energy and Buildings*, 36(10):981–986, 2004.

[180] R. Kosonen and F. Tan. Assessment of productivity loss in air-conditioned buildings using PMV index. *Energy and Buildings*, 36(10):987–993, 2004.

[181] Raimo Niemelä, Jorma Railio, Mika Hannula, Sari Rautio, and Kari Reijula. Assessing the effect of indoor environment on productivity. In *Clima 2000*, Napoli, Italy, 2001.

[182] Raimo Niemelä, Mika Hannula, Sari Rautio, Kari Reijula, and Jorma Railio. The effect of indoor air temperature on labour productivity in call centers – a case study. *Energy and Buildings*, 34(8):759–764, 2002.

[183] Clifford C. Federspiel, G. Liu, M. Lahiff, D. Faulkner, D. L. Dibartolomeo, W. J. Fisk, P. N. Price, and D. P. Sullivan. Worker performance and ventilation: Analyses of individual data for call-center workers. In *Indoor Air*, pages 796–801, 2002.

[184] J. Link and R. Pepler. Associated fluctuations in daily temperature, productivity and absenteeism. *ASHRAE Transactions*, 76(2):326–337, 1970.

[185] G. Meese, R. Kok, M. Lewis, and D. Wyon. A laboratory study of the effects of moderate thermal stress on the performance of factory workers. *Ergonomics*, 27(1):19–43, 1984.

[186] Kasper L. Jensen, Jørn Toftum, and Peter Friis-Hansen. A Bayesian Network approach to the evaluation of building design and its consequences for employee performance and operational costs. *Building and Environment*, 44:456–462, 2009.

[187] Fergus Nicol, Mike Wilson, and Cecilia Chiancarella. Using field measurements of desktop illuminance in European offices to investigate its dependence on outdoor conditions and its effect on occupant satisfaction, and the use of lights and blinds. *Energy and Buildings*, 38(7):802–813, 2006.

[188] Henri Juslén, Marius Wouters, and Ariadne Tenner. The influence of controllable task-lighting on productivity: a field study in a factory. *Applied Ergonomics*, 38(1):39–44, 2007.

Frédéric HALDI
9, rue de l'Industrie
1005 Lausanne
Switzerland
E-mail: frederic.haldi@epfl.ch

Age 30
Married
Swiss nationality

* * *

Education

2005 – 2009	**PhD** at EPFL (Solar Energy and Building Physics Laboratory), thesis on the development of probabilistic models for the prediction of human behaviour, to improve building energy performance simulation methods
1998 – 2003	**Master in physics** at Geneva University (Department of Theoretical Physics), thesis in the domain of fluid mechanics
1994 – 1998	**High school certificate** (Maturité fédérale) with scientific orientation at Collège Claparède (Geneva), with distinction

Professional experience

Since 2005
(4 years)

EPFL, Solar Energy & Building Physics Laboratory (Lausanne, Switzerland)
Research and teaching assistant
- Development of probabilistic models for the prediction of actions performed by buildings' occupants (windows, shading devices) and their associated impact on building energy performance (air flows, solar gains)
- Advances in the methods for the assessment of indoor environment quality in naturally ventilated buildings
- Formulation of these models for application in a dynamic simulation program for the prediction of urban energy flows
- Development of a model to predict the risk of overheating in buildings
- Teaching of building physics to undergraduate students in architecture

2004 – 2005
(1 year 3 months)

State Department of Education (Geneva, Switzerland)
Teacher
Teaching of English, German, Latin, Mathematics and Physics in several secondary schools

2004
(5 months)

Sustainable Energy Ireland (Cork, Ireland)
Technical consultant for *Renewable Energy Information Office*
- Programming of calculation tools to assess the potential of biomass as an energy source in Ireland
- Contribution to design of a digital information tool on renewable energies for information of the public and enterprises
- Participation to projects linked with energy management and building physics

2000-2003
(3 years)

Collège Claparède (Geneva, Switzerland) – Physics laboratory assistant
Preparation of experiments and theoretical developments in collaboration with students

Skills

Building physics
Building thermal simulation, energy performance of buildings, SIA 380/1 and Minergie norms, indoor environment quality

Statistical modelling
Random processes, behavioural modelling, discrete choice analysis, survival analysis, linear and non-linear models, design of experiments, financial models

General physics
Fluid mechanics, Navier-Stokes equations, thermodynamics and statistical physics

Computer science
Operating systems : Windows, UNIX, Mac OS
Office : Proficient in Word, Excel, Access, Power Point, Photoshop, Illustrator, LaTeX
Scientific software : Proficient in Matlab, Maple, R and S-PLUS
Programming languages : Proficient in Turbo Pascal and Delphi, familiar with C++
Building simulation : Proficient in ESP-r, Lesosai and PHPP, familiar with EnergyPlus and ECOTECT

Languages

French	mother tongue, excellent writing skills
English	fluent (C2), five months professional stay in Ireland (2004), good writing skills
German	conversant (B2)
Russian	fluent (C1), fours stays of one month in Russia and Ukraine (1997-2004), TRKI-1 diploma of Moscow State University (2002)
Spanish	conversant (B2), one month stay in Argentina (2003)
Portuguese	conversant (B2), two months stay in Brazil (2003)
Serbo-croatian	notions (B1)

Academic awards

2009	Best Paper Award from journal *Building and Environment*
2009	Best Student Paper Award of International Building Performance Simulation Association (IBPSA) Conference

List of publications

Refereed journal articles

F. Haldi, D. Robinson, Adaptive actions on shading devices in response to local visual stimuli, Journal of Building Performance Simulation, 2010, Article in press

F. Haldi, D. Robinson, Interactions with window openings by office occupants, Building and Environment, Volume 44, Issue 12, December 2009, Pages 2378-2395

D. Robinson, F. Haldi, An integrated adaptive model for overheating risk prediction, Journal of Building Performance Simulation, Volume 1, Issue 1, March 2008, Pages 43-55

F. Haldi, D. Robinson, On the behaviour and adaptation of office occupants, Building and Environment, Volume 43, Issue 12, December 2008, Pages 2163-2177

D. Robinson, F. Haldi, Model to predict overheating risk based on an electrical capacitor analogy, Energy and Buildings, Volume 40, Issue 7, July 2008, Pages 1240-1245

F. Haldi, P. Wittwer, Leading order down-stream asymptotics of non-symmetric stationary Navier-Stokes flows in two dimensions, Journal of Mathematical Fluid Mechanics, Volume 7 (2005) 611-648

Refereed conference articles

P. Renaud, J. Hars, C. Piot-Ziegler, D. Robinson, F. Haldi, C.-A. Roulet, Explosion of energy demand for air cooling in summer: Perspectives and solutions (EEDACS), In Proc. CISBAT 2009, p. 175-180

D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, From the neighbourhood to the city: Resource flow modelling for urban sustainability, In Proc. CISBAT 2009, p. 445-450

D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning, In Proc. 11th Int. IBPSA Conf: Building Simulation 2009, pages 1083-1090, Glasgow, United Kingdom

F. Haldi, D. Robinson, A comprehensive stochastic model of window usage: Theory and validation, In Proc. 11th Int. IBPSA Conf: Building Simulation 2009, pages 545-552, Glasgow, United Kingdom

F. Haldi, D. Robinson, A comprehensive stochastic model of blind usage: Theory and validation, In Proc. 11th Int. IBPSA Conf: Building Simulation 2009, pages 529-536, Glasgow, United Kingdom

F. Haldi, D. Robinson, A comparison of alternative approaches for the modelling of window opening and closing behaviour, In Proc. Air Conditioning and the Low Carbon Cooling Challenge, Windsor, United Kingdom, 2008

D. Robinson, C. Giller, F. Haldi, H. Fei, J. Kämpf and A. Kostro, Towards comprehensive simulation and optimisation for more sustainable urban design, In 15. Schweizerisches Status-Seminar 2008, pages 489-496, Zürich

F. Haldi, D. Robinson, An integrated adaptive model of overheating risk, In 15. Schweizerisches Status-Seminar 2008, pages 1-8, Zürich

F. Haldi, D. Robinson, Stochastic / Probabilistic modelling of multiple adaptive processes: some subtle complexities, In Proc. e-Sim 2008 Conference, Québec, Canada, 2008

D. Robinson, F. Haldi, An integrated adaptive model for overheating risk prediction, In Proc. 10th Int. IBPSA Conf: Building Simulation 2007, pages 745-750, Beijing, China

F. Haldi, D. Robinson, Representing Behaviour and Adaptation of Office Occupants in Building Simulation, In Proc. CISBAT 2007, pages 319-324, Lausanne, Switzerland, 2007