On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption

THÈSE N^{O} 6881 (2016)

PRÉSENTÉE LE 31 AOÛT 2016

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE INTERDISCIPLINAIRE DE PERFORMANCE INTÉGRÉE AU PROJET
PROGRAMME DOCTORAL EN GÉNIE CIVIL ET ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Parag RASTOGI

acceptée sur proposition du jury:

Prof. A. Rinaldo, président du jury
Prof. M. Andersen, directrice de thèse
Prof. J. Hensen, rapporteur
Dr S. Natarajan, rapporteur
Prof. A. Davison, rapporteur

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Suisse
2016

In memory of my dear grandfather, nanaji,
who didn't live to see this finished.

To naniji, ma, papa, and didi.

DON'T PANIC

Acknowledgements

… But 'tis a common proof,
That lowliness is young ambition's ladder,
Whereto the climber-upward turns his face;
But when he once attains the upmost round.
He then unto the ladder turns his back,
Looks in the clouds, scorning the base degrees
By which he did ascend.

The Life and Death of Julius Caesar, Act 2, Scene 1.
William Shakespeare, circa 1599[1]

Given that scientific writing is terminally dry and often deliberately abstruse, I will write this pseudo-section with a far freer vocabulary and style than the rest of the thesis. It's not that I'm not serious about acknowledging the people who helped me reach here, it's just that I'd prefer to do so without resorting to the passive voice and semi-colons. I will skip the 'thank you' for each sentence/person, because the whole point of this write-up is to say a big THANK YOU! The people who've contributed in ways big and small are listed here in no particular order[2]. And to those whom I've missed, my apologies. Know that your work and input has influenced me in ways I cannot describe. There really is no retreating from a charge of pedantry after quoting Shakespeare. However, just because this thesis is slightly pretentious in its results and language does not mean that I scorn the base degrees by which I did ascend, like Caesar allegedly did. So, friends, family, scientists, thank you: you believed in me when I didn't believe in myself.

This thesis could not have been realised without the support, attention, cross-examination, criticism, and generosity of Marilyne Andersen (the first triumvir). You brought me to Switzerland, gave me all the freedom and backing I could have asked for, and pushed me to be a better academic and researcher.

Quite literally, none of this would have been possible without my parents, Anita (Rosy) and Deepak (Rusty). You are the foundation on which all of this is built. Apart from carefully

[1] *shakespeare.mit.edu/julius_caesar*
[2] OK, there is *some* order… but it is more categorical than ordinal and should be interpreted loosely.

nurturing this difficult child, you gave me all the books, culture, values, and opportunities a boy could possibly want. My dear sister, Garima (Indu), who has been my chief ally, rival, booster, school-ground-defender, and all-round brother and sister rolled into one. My grandparents, the Raj-Aryas and the Rastogis, who encouraged me to dream.

Prof. Davison, master teacher, giver of advice, infinitely patient, and (acceptable) cups of tea. Your insights immeasurably strengthened the structure of this thesis, and your reviews made it far more comprehensible. Prof. Hensen and Prof. Natarajan, for bravely accepting to read the thesis on the basis of a few meetings and presentations, and for sharing your experience and knowledge with me. Prof. Rinaldo, both for agreeing to chair my committee, and for your passion for Fourier Analysis and tough mathematics, the influence of which is clearly perceptible in this document. Prof. Bierlaire, whose class was equal parts enjoyable, challenging, and useful, and whose supportive stewardship of the doctoral school I certainly benefited from. EPFL, the great mother ship, where this thesis was undertaken, and of whose facilities and jogging paths I heartily availed myself.

My dear officemates, colleagues, friends, critics, students, and teachers: Shevy, Maria, Emilie, Giorgia, Kynthia, Minu, Giuseppe, Mandana, Sergi, Luisa, Jan, Shady, Massi, Martine T, Martine L, Joelle, Peter, Judith, Loic, Thomas M, Sharon, Lorenzo, Boris (the second triumvir). Without your constant help, feedback, resourcefulness, and willingness to contribute, I wouldn't have survived this[3]. Those who put up with me as a supervisor and instructor: Soenke, Margaux, Michalis, Koven, Guillaume, Helene, and the several batches of ENAC students left wondering how energy simulation could possibly be considered 'fun'. I learnt more from you than I could ever hope to give back.

The invaluable friends, cultivated over conferences, meetings, casual encounters, ski trips, and beer sessions, in Switzerland and elsewhere: Mikael, Roel, George M, Emti, Carlos, Jamie, Ciara B, Chiara M, Philippe, Dimitri M, Michi, Hendrik, Stefano, Alex B-S, Sagar, Jignesh, Bharat K, Alfonso, Fabio F, Mikko, Ben M, Ralph E, Govinda, Naomi, Mike D, Pierre L, Vahid N, Christi, Leo, Michelle, Catherine, Prof. de Wilde, Dr. Joe Huang, Dr. Crawley. Thank you for letting me pick your brains, use your facilities, block up your computers, and nearly ruin my ligaments/liver, in the pursuit of *knowledge*!

Professors and mentors from Purdue, who introduced me to the world of engineering: Thanos, Panagiota, Jim Braun, Travis, Judy Liu, Becky Hull, Alex Gluhovsky, and the School of Civil Engineering and College of Engineering. My teachers from school, without whose efforts I wouldn't be able to read or add properly, or consider science a worthwhile life goal: Mr Singh (DVS), Mr Chamola (BKC), Ms Chaturvedi (PCH), Dr Farooqui (MHF), Ms Kuthiala (STK), Ms Dutta (PDT), Mr Pandey (VNP), Mr Ahuja (BLA), Mr Yusufji (SAY), Dr Bajpai (KPB), Ms Bhattacharya (PKB), Dr Joshi (MCJ), and Ms Nirupama Goel and Ms Abha Naithani from Welham. The lifelong friends made in school, who I can always turn to, for anything – Ohri, Keswani, Sharma, Bector, Singhi, Kuthiala, Jagabanta, and, especially, Aditi.

[3] And my graphs/presentations would look even worse than they do now...

This thesis is full of interesting bits and pieces that could be useful to others in the future, but also flaws and oversights. It is built on the work of far more talented women and men, both scholars in the field and the foot-soldiers of science, going back to the first human being who just wanted to find out what would happen if two rocks were struck together. My work was made a lot easier and smoother by the thousands of silently toiling geeks and nerds who write software, code snippets, and detailed answers on-line without any material rewards, but whose efforts probably underpin the internet. It would be impractical to name everybody, but their help is gratefully acknowledged. For every idea, technique, or algorithm that sort-of, kind-of, works in this thesis, there are dozens of ideas that failed. There are at least 10 versions of every script presented here and online, but if I had not failed with dozens of ideas, I would not have found the ones that worked. Happy reading, and remember: DON'T PANIC!

Lausanne, 17th August 2016 P R

Pramânasiddhântaviruddham atra
Yatkiñciduktam matimândyadosât
Mâtsaryyam utsâryya tadâryyacittâh
Prasâdam âdhâya vis'odhayantu.

May the noble-minded scholars,
instead of cherishing ill feeling,
kindly correct whatever errors have been here committed
through the dullness of my intellect
in the way of wrong interpretations and misstatements.

Hemachandra
quoted and paraphrased by Surendranath Dasgupta in,
A History of Indian Philosophy, Vol.1 (1922).

Abstract

Building simulation requires a large number of uncertain inputs and parameters. These include quantities that may be known with reasonable confidence, like the thermal properties of materials and building dimensions, but also inputs whose correct values cannot be known with absolute certainty, notably weather and occupancy. Building simulation is commonly used to estimate the impact of design decisions on indoor conditions to enable relative comparisons. A simulation run is not, strictly, a prediction. Since the parameters and calculations are approximations of real-world phenomena and materials, the exercise is essentially uncertain. Regardless of whether simulation is interpreted as a prediction or an approximation indicative of average behaviour, including explicit bounds of uncertainty is more informative for a decision-maker than a single point estimate.

Climate or weather as input to building simulation is the dominant theme of this thesis. Current practice calls for the use of a typical weather file to evaluate design choices based on energy consumption. The typical year file is intended to represent the mean climate, and therefore mean energy usage: it does not represent the range of impact on the final indoor environment. That is, a confidence or variability interval about the mean response cannot be calculated using a typical year file.

This thesis presents results for two related but independent proposals for sensitivity and uncertainty analyses in building simulation, particularly to weather. The first is a novel, generalisable procedure for generating synthetic weather data to carry out a Monte Carlo experiment with a building simulation model. The second is a technique for training emulators or response surfaces to rapidly obtain estimates of performance outputs from simulation models, using Gaussian Process regression on small training data sets. The two parts, together and separately, enable the quantification of the lack of knowledge about an input, and the impact of this uncertainty on the final results.

The synthetic weather time series developed are an ensemble of realistic hourly data whose mean statistical characteristics are close to the typical year used to generate them. The procedures developed are generalisable with minimal expert input. We avoid presenting a unified model for all climates, leaving some tuning parameters like the extent of correlation, and the unknown coefficients of stationary time series models, to be calculated empirically (based on the typical file of a given climate). The emulators are created using regression, comparing the performance of classical parametric regression with a non-linear technique based on Gaussian random processes. The issue with representing a highly non-linear and non-smooth process like building simulation with classical regression is that the models are only reliable within very restricted sampling domains. We were able to overcome this with Gaussian Process regression. Our proposal trains reliable models on small samples, reducing the computational burden, and gives an explicit estimate of the uncertainty for a prediction, since the response at any sampled point is modelled as a Normally-distributed random process. Once again, we avoid a unified emulator or regression model because the response from one building (defined by its geometry and usage in this case) is not necessarily an appropriate description of the response of another.

This work is a step towards practical tools for the use of building simulation in a stochastic paradigm. Both elements of the thesis contribute toward explicitly estimating the uncertainty in the results of building simulation, using empirical or data-driven techniques. The types of the time series and emulator models are general enough to work on any climate or building, with parameters obtained from the simulated/typical sample at hand, but the importance of different aspects and the nature of a building's response are determined uniquely (i.e., parameter values). The work is easily extensible to the analysis of the sensitivity of a building, or groups of buildings, to any inputs. The concepts proposed in this thesis may also be used for stochastic optimisation and models to predict performance metrics other than the annual sum of energy.

Keywords:
adaptation, building simulation, climate change, climate resilience, climate risk, energy efficiency, sensitivity analysis, stochastic simulation, weather generator, uncertainty analysis

Résumé

Sur la sensibilité des bâtiments au climat : l'interaction des conditions climatiques et de l'enveloppe des bâtiments sur la détermination de leur consommation d'énergie future

La simulation des bâtiments nécessite une bonne connaissance d'une multitude de données d'entrée, de facteurs incertains et de paramètres. Ceux-ci incluent des quantités qui peuvent être connues avec suffisamment de certitude, comme les propriétés thermiques des matériaux et les dimensions du bâtiment, mais aussi de certains autres paramètres qui ne peuvent pas être définis de façon satisfaisante, notamment la météo et l'utilisation effective du bâtiment. La simulation est alors plutôt un outil pour estimer l'impact des décisions de conception sur les conditions intérieures, afin de permettre des comparaisons relatives et non une prédiction. Les paramètres des calculs sont des approximations des phénomènes et des matériaux réels, donc l'exercice est fondamentalement incertain. Les résultats de la simulation devraient être interprétés comme des prédictions ou des indications de performance moyenne. Ceci inclut des déclarations explicites sur l'incertitude d'un certain résultat, plus communément exprimé comme intervalle de confiance autour d'une réponse moyenne.

Le thème principal de cette thèse est la considération du climat et de la météorologie comme données d'entrée dans la simulation de la consommation d'énergie d'un bâtiment. L'usage se veut d'utiliser des données typiques pour évaluer le design. Le fichier climatique annuel représente le climat moyen, et ainsi l'énergie moyenne utilisée, et non l'ensemble des influences du climat sur l'environnement intérieur. En d'autres mots, un intervalle de confiance ne peut pas être calculé en utilisant un seul fichier climatique annuel.

Cette thèse présente les résultats de deux approches liées mais indépendantes pour faire des analyses de sensibilité et d'incertitude. La première est une nouvelle procédure généralisable pour produire des données météorologiques synthétiques afin d'effectuer une simulation de Monte Carlo d'un modèle de simulation d'un bâtiment. Le second est une technique pour former des émulateurs ou « response surfaces » à partir d'un nombre réduit de données, afin d'obtenir rapidement des estimations de

performance. Les deux parties abordent, ensemble et séparément, l'enjeu de souligner explicitement le manque de connaissances sur une certaine donnée d'entrée ainsi que l'impact de ce manque sur les résultats finaux.
Les données météorologiques synthétiques sont un ensemble de données horaires réalistes dont les statistiques caractéristiques moyennes sont proches de l'année météorologique typique utilisée pour les générer. Les procédures utilisées sont généralisables, suivant une contribution minimale de la part d'experts pour assurer un contrôle de la qualité. Les émulateurs sont développés en utilisant des techniques de régression : la performance de régression classique paramétrique est comparée à une technique non-linéaire basée sur des processus aléatoires gaussiens (*Gaussian Process regression*). Le problème lié à la représentation d'un processus hautement non-linéaire et non-lisse, comme la simulation d'un bâtiment à travers une régression classique, est que les prédictions ne sont fiables que dans des domaines d'échantillonnage très restreints. La *Gaussian Process regression* produit des modèles fiables même sur de petits échantillons, réduisant ainsi la charge de calcul et donnant une estimation explicite de l'incertitude pour une prédiction, puisque la réponse est modélisée comme un processus aléatoire Gaussien. Encore une fois, nous évitons un modèle unifié car la réponse d'un bâtiment (défini par sa géométrie et son utilisation dans ce cas) n'est pas nécessairement représentative de la réponse d'un autre bâtiment.
Le travail présenté dans cette thèse est une étape vers des outils pratiques pour l'utilisation de la simulation des bâtiments dans un paradigme stochastique. Les deux éléments de la thèse contribuent à estimer l'incertitude dans les résultats de simulation. Les techniques présentées sont empiriques mais leurs structures sont constantes et robustes. Les formes des modèles de série temporelle et d'émulation proposés ici sont suffisamment générales pour fonctionner sur tous les climats et bâtiments, mais l'importance des différents aspects et la nature de la réponse d'un bâtiment sont déterminés de manière spécifique. Le travail est facilement extensible pour l'analyse de sensibilité d'un bâtiment ou d'un ensemble de bâtiments, à tout type de données d'entrée. Les concepts proposés dans cette thèse peuvent aussi être utilisés pour faire une optimisation stochastique et de concevoir des modèles afin de prédire d'autre métriques de performance autre que la somme annuelle d'énergie.

Contents

Acknowledgements **iii**

Abstract (English) **vii**

Résumé (Français) **ix**

1 Introduction **1**
1.1 A Provocation 1
1.2 This Thesis: Novelty and Organisation 5
1.2.1 Contribution 5
1.2.2 Thesis Structure 7
1.2.3 Terms and Terminology 8
1.3 Climate and Buildings 12
1.3.1 History, and Business As Usual 12
1.3.2 Climate Change 14
1.4 Energy-Conscious Building Design in an Uncertain Climate 15
1.4.1 Comfort and Expectations 16
1.4.2 The Role of Building Performance Simulation 17
1.5 Quantifying Uncertainty, Sensitivity, or Both? 18
1.5.1 Does Simulation Imply Prediction? 18
1.5.2 Why Abandon Certainty? 21
1.5.3 Numerical vs Analytical Approaches 22
1.5.4 Uncertainty Analysis 23
1.5.5 Sensitivity Analysis 24
1.6 Simulation Inputs: Perspective and Interpretation 25
1.6.1 Weather Inputs 25
1.6.2 Building Properties 27

2 State of the Art **29**
2.1 Climate Classification and Characterisation 30

2.1.1 Historical Development . . . 30
2.1.2 Contemporary Work . . . 33
2.1.2.1 ASHRAE Climate Zones . . . 33
2.1.2.2 Climate Severity Index . . . 34
2.1.2.3 Other Classification Systems . . . 35
2.2 Future Climate, Climate Change, and Buildings . . . 36
2.2.1 Future Climate Inputs . . . 37
2.2.2 Impact of Climate Change . . . 41
2.3 Simulation for Energy-Conscious Design . . . 43
2.3.1 Comfort Models for Simulation: A Brief Overview . . . 44
2.3.2 Buildings as Systems to be Simulated . . . 45
2.3.3 Simulation: Usability Issues . . . 46
2.3.4 Using Simulation for Early- and Late-Stage Design . . . 49
2.3.5 Optimising for Energy and Comfort . . . 50
2.4 Uncertainty and Sensitivity in Simulation . . . 52
2.4.1 Errors vs Uncertainty . . . 55
2.4.2 Types of Uncertainty . . . 55
2.4.2.1 Aleatory . . . 55
2.4.2.2 Epistemic . . . 56
2.4.3 Uncertainty Propagation . . . 57
2.4.3.1 Internal Methods: Stochastic Models . . . 59
2.4.3.2 External Methods: Random Inputs . . . 60
2.4.4 Sensitivity Analyses in Building Simulation . . . 62
2.4.5 Issues: Speed and Complexity . . . 63
2.5 Simplification and Emulation for Speed . . . 64
2.5.1 Simplified Physical Models . . . 65
2.5.2 Databases and Regression . . . 67
2.6 Weather input for Simulation . . . 69
2.6.1 Typical or Reference Weather Years . . . 70
2.6.1.1 Finkelstein-Schafer Years . . . 71
2.6.1.2 ASHRAE Test Reference Years . . . 73
2.6.1.3 Extreme Years . . . 73
2.6.1.4 Typical Principal Component Years . . . 74
2.6.1.5 Others . . . 75
2.6.2 The Problem with Typical Data . . . 77
2.6.3 Weather Generators and Synthetic Data . . . 79
2.7 Uncertainty in and due to Weather Inputs . . . 85
2.7.1 Temporal Uncertainty: Climatic Volatility and Climate Change . . . 86

2.7.2 Spatial Uncertainty: Natural and Human Factors 88
2.8 Summary 91

3 Synthetic Weather Inputs for Building Simulation 93
3.1 General Approach 93
3.2 The Synthetic Generation Procedure 94
3.3 Splitting the Time Series into Deterministic and Random Components 96
3.4 Fourier Fitting To Remove Seasonal Trends 98
3.4.1 Modifications for Climate Change Forecasts 99
3.4.2 Characteristics of the Residuals 101
3.5 Stationary Time Series Models for Random Components 105
3.5.1 Correlation Estimation 106
3.5.2 Selecting a Model 106
3.5.3 The Selected SARMA Model 107
3.5.4 Characteristics of the Residuals 109
3.6 Introducing Noise for Variation 111
3.6.1 Resampling and Subsampling 111
3.6.2 Simulating the SARMA model 114
3.7 Recombining Random and Deterministic Components 116
3.8 Post-Processing the Synthetic Series 117
3.9 Special Treatment for Solar Quantities 118
3.10 Examining the Synthetic Weather Series 120
3.10.1 Probability Distributions and Dispersion 121
3.10.2 Sequences 125
3.10.3 Cross-Correlation 126
3.10.4 Measures of Shape 129
3.10.5 Random Files with Climate Change 130
3.11 Simulations with Synthetic Files: An Example 131
3.12 Limitations 133
3.12.1 Episodes and Extremes on Demand 133
3.12.2 Expert Input 134
3.12.3 Choice of Source File 135
3.12.4 Number of Runs 136
3.13 Synthetic Generation: Summary 139

4 Emulators for Uncertainty and Sensitivity Analyses 141
4.1 General Approach 141
4.2 Inputs and Outputs 143
4.2.1 Case Studies 143

4.2.2 Choosing Inputs: Orthogonality and Dependence 145
4.2.3 Outputs: Distributions and Smoothness 149
4.2.4 Transformation and Scaling of Data 152
4.2.5 The Fitting Procedure 153
4.2.6 Prediction Error Analysis and Model Selection 154
4.3 Step One: Classical Emulators 156
4.3.1 Classical Linear Models 156
4.3.1.1 Normal Linear Models 156
4.3.1.2 Generalised Linear Models and Linear Mixed Models . 157
4.3.2 Analysis of Variance (ANOVA) 158
4.3.3 Results and Discussion: Linear Models 158
4.3.3.1 Guide to Linear Regression Plots 158
4.3.3.2 Prediction and Error Plots 159
4.3.4 What about Non-Linear Fits? 165
4.4 Step Two: Probabilistic Emulators 165
4.4.1 Gaussian Processes and Regression 166
4.4.1.1 A Very Brief Introduction 166
4.4.1.2 K-fold Cross-Validation 167
4.4.1.3 Choice of Kernels 169
4.4.2 Results and Discussions: Gaussian Process Models 170
4.4.2.1 Guide to GP Plots 170
4.4.2.2 Plots: Squared Exponential Kernel 171
4.5 Evolution of Error: Transversal Plots 185
4.6 Emulators: Summary 188

5 Conclusion **191**
5.1 Contribution 191
5.2 Workflow Summary 194
5.3 Scope and Applications 196
5.3.1 Random Simulation as Design-Assistance 196
5.3.2 Renovation Strategies and the Performance Gap 197
5.3.3 Comfort Assessment 198
5.3.4 Model Predictive Controls 199
5.3.5 Modelling Stocks, Grids, and Renewables 200
5.3.6 Stochastic or Robust Optimisation with Uncertain Inputs 201
5.3.7 Regression of Internal Temperatures 201
5.3.8 Random Occupancy Modelling 202
5.4 Outlook 202

A Additional Results and Concepts: Synthetic Weather **207**
A.1 Notes on Implementation 207
A.2 Fourier Series and Fitting 208
A.3 Stationary Time Series Models 209
A.3.1 Auto-Regressive (AR) Models 211
A.3.2 Moving Average (MA) Models 211
A.3.3 Auto-Regressive Moving Average (ARMA) Models 212
A.3.4 Homoscedasticity and Conditional Variance Models 212
A.3.5 Script Snippet for Solar Series Bootstrap 213
A.4 Additional Results 214
A.5 Weather Data Sources 219

B Regression: Additional Concepts and Details **221**
B.1 Confidence Intervals 221
B.2 More Gaussian Processes 223
B.2.1 GP Kernels: Additional Details 223
B.2.2 Implementation 225
B.2.2.1 Fit Method 225
B.2.2.2 Predict Method 226
B.2.2.3 Basis Function 226
B.2.2.4 Standardization and Regularization 226
B.2.3 Results: SQE Kernel 227
B.2.4 Results: ARD Kernel 236
B.3 Regression Inputs: Additional Discussion and Concepts 245
B.3.1 List of Initial Inputs 245
B.3.2 Dependence and Orthogonality 247
B.3.2.1 Checking Correlation 247
B.3.2.2 Principal Component Analysis 250
B.3.3 Scaling Options for Inputs and Outputs 251
B.4 Simulation Details 252

Bibiliography **263**

Glossary **287**

Curriculum Vitae **301**

1 Introduction

... the universe was full of ignorance all around and
the scientist panned through it like a prospector
crouched over a mountain stream,
looking for the gold of knowledge
among the gravel of unreason,
the sand of uncertainty, and
the little whiskery eight-legged swimming things of superstition.

Terry Pratchett, *Witches Abroad*

1.1 A Provocation

In recent years, energy and sustainability issues have vied with access and cost on the top of the agenda for the built environment. Put simply, the challenge is to ensure that every human being has access to a comfortable and safe shelter, without destroying the environment. And, to paraphrase the Brundtland Commission (World Commission on Environment and Development 1987), we should have been done with this yesterday, without exhausting the resources we will need tomorrow.

It is not surprising that much of the initial research or thought about indoor conditions went into improving indoor lighting and ventilation. The industrial revolution, with its "*dark Satanic mills*[1]" and filthy cities, mostly concentrated in northern Europe, had a profound impact on the living conditions of the urban populace, much of which

[1] William Blake was probably not thinking of building simulation when he wrote this though, c. 1808.

had formerly been the rural populace. Since then, admirable progress has been made in the condition of the built environment, at least in the industrialised world. In an unhappy twist, the definition of 'modernity' widely adopted by developing nations today creates buildings that do not give much importance to indoor environmental quality. While the resulting indoor environments are not as bad as those from the early decades of the industrial revolution, the pervasive idea of using energy to compensate for inappropriate design has not gone away. We now have all sorts of clever ways of measuring the impacts of indoor environmental conditions on human health and productivity. Somewhat ironically, much of it boils down to the same prescriptions as a hundred years before the industrial revolution: plenty of fresh warm/cool air, exercise, and avoidance of toxic substances. Most societies regard access to a safe and comfortable indoor environment, at home and work, as a necessity, if not a right.

Lack of access to warm or cool interiors in extreme situations can be fatal or, at the very least, exhausting; and, anybody who has spent time in a tropical country or, more recently, in a city even in a temperate country, will swear by the benefits of air conditioning and filtration. In the modern world, therefore, it is not hard to see a comfortable indoor environment as a *necessity*. Yet, the definition of this necessity keeps expanding. As Ackermann (2010) argues persuasively in her book, the "culture of cool", and its accompanying energy crisis, arise from largely *manufactured needs*. Infrastructure does not just meet demand, it may also create it. It is almost too easy to point out the excesses of glass towers in the desert heat, contemporary travesties of Corbusier's *Mur Neutralisant* and *Respiration Exacte*, enabled by space-age materials and cheap fuel. However, these buildings arise from a certain set of economic and social factors, and their damage to the environment is more symbolic than statistically significant. In a perverse way, these buildings are very efficient and high-performing. In the quest for ever-more efficiency and higher performance, a culture that values hermetic bubbles over the wholesale improvement of the built environment has taken hold. It will be far more difficult to tackle the rising demand for cooling and heating from a billion new dwellings packed into ever-denser cities if we do not address this cultural conditioning. Therefore, it befits us to consider the *simpler, even primitive, aspects of the problems we are seeking to solve*.[2]

We contend that working *with* the climate rather than *against* it, to create the relatively small range of thermal conditions that humans prefer, is a far more sound course of action – environmentally, socially, and economically – than any number of efficient fan-coils. Accepting a lack of control and thermal variability to create delight, rather

[2] It is difficult to point where exactly, but the discussion in this chapter has been influenced by Banham (1984), so we will just generally acknowledge his work here.

than just to avoid discomfort, may lead to simple buildings, but the occupants of these buildings may learn to adapt far more. A changing climate and more frequent extremes will necessitate the use of systems, and there are certain climates which cannot be tackled with clever design alone, but Heating, Ventilation, and Air Conditioning (HVAC) systems used as *back-ups* rather than *defaults* are certainly a desirable outcome. The philosophical bent of this thesis is not to suggest that the solution to the energy crisis is a large scale roll-back of the (thermal) comforts we have come to take for granted.

This thesis takes the optimistic and somewhat self-serving view that the climate is a *very important* factor in building design, second only to the wishes of the user. Naturally, nothing can be constructed without a budget and availability of materials. All else being the same, though, climate and usage should guide everything from site selection through massing to the number of panes in each window. With the advent of mechanised climate control and inexpensive energy, this order of preference has been turned on its head somewhat. However, energy is not limitless (yet), and some of the by-products of energy production – pollution, environmental degradation, and Green House Gas (GHG) emissions – tend to undermine the collective efforts of humanity to improve its living conditions. Our view is optimistic in the sense that we expect new generations of architects and engineers to regard bioclimatic (read: energy-conscious) design as the norm. Ticking check-lists of prescriptive norms about insulation and window conductance can hardly be the basis for 21st century building design.

This work is not an attempt to provide concrete rules and guidance for bioclimatic architecture. Rather, we focus on the mathematical tools to provide additional knowledge of certain factors and their interactions, and to quantify our lack of knowledge about these quantities. Specifically, this thesis describes methods to analyse the sensitivity of indoor environmental conditions to certain properties of the climate and building envelope. Where permitted by the mathematics of the problem, we quantify uncertainty about the values of these exogenous (lit. 'outside of the system') variables, climate and envelope, on the expected energy need for space heating and cooling. The procedures described in this thesis should not be interpreted as *value judgements*, but as *epistemological quantifiers*: they are not meant to show what is good or bad, they are only meant to help quantify one's knowledge, or lack of it, about the inputs, outputs, and their relationships. Human judgement does not need to be replaced by ever-smarter algorithms or tools, or yet another set of guidelines. What we aim to show is how much impact *not knowing* something (e.g., future weather) has on something else one is trying to control (e.g., indoor comfort, energy use).

We begin with a prophetic warning from Douglas Adams.

THE Great Ventilation and Telephone Riots of SrDt 3454 had started off as just a lot of hot air.

Hot air was, of course, the problem that ventilation was supposed to solve and generally it had solved the problem reasonably well up to the point that someone invented air-conditioning, which solved the problem far more throbbingly. And that was all well and good, provided you could stand the noise and the dribbling until someone else came up with something even sexier and smarter than air-conditioning, which was called in-building climate control. Now this was quite something. The major differences from just ordinary air-conditioning were that it was thrillingly more expensive, and involved a huge amount of sophisticated measuring and regulating equipment which was far better at knowing, moment by moment, what kind of air people wanted to breathe than mere people did. It also meant that, to be sure that mere people didn't muck up the sophisticated calculations which the system was making on their behalf, all the windows in the buildings were built sealed shut. This is true. While the systems were being installed, a number of the people who were going to work in the buildings found themselves having conversations with Breathe-O-Smart systems fitters which went something like this:

"But what if we want to have the windows open?"

"You won't want to have the windows open with new Breathe-O-Smart."

"Yes, but supposing we just wanted to have them open even for a little bit?"

"You won't want to have them open even for a little bit. The new Breathe-O-Smart system will see to that."

"Hmmm..." "Enjoy Breathe-O-Smart!"

"Okay, so what if the Breathe-O-Smart breaks down or goes wrong or something??"

"Ah! One of the smartest features of the Breathe-O-Smart is that it cannot possibly go wrong. So. No worries on that score. Enjoy your breathing now, and have a nice day."

...

Major heat waves started to coincide, with almost magical precision, with major failures of the Breathe-O-Smart systems. To begin with, this merely caused simmering resentment and only a few deaths from asphyxiation.

Douglas Adams, *Mostly Harmless* (1992).

1.2 This Thesis: Novelty and Organisation

...it is well known that a vital ingredient of success
is not knowing that what you're attempting can't be done.
A person ignorant of the possibility of failure can be
a half-brick in the path of the bicycle of history.

Terry Pratchett, *Equal Rites*

We begin by discussing our approach – what is unique about it, why it is relevant, and how it can be of use to simulation practitioners – and what the reader can expect in the rest of this thesis. This is discussed again in the final chapter (Section 5.3).

1.2.1 Contribution

As we have discussed elsewhere, modern Building Performance Simulation (BPS) tools do not explicitly solve core differential equations stochastically, with few exceptions. Rather, they rely on multiple runs with random inputs to quantify uncertainty or sensitivity. That is, unless one is explicitly modelling changing material properties, or randomly seeding a ray-tracing algorithm, every simulation run of a given building model can be expected to give exactly the same output. The approach advocated in this thesis changes this paradigm: we ask the user to *not* expect the same answer with every simulation. Instead, every simulation may be interpreted as an experimental run, and the sensitivity of the system being simulated determines the variation in output that can be expected from uncertain inputs.

The intention of our work related to the creation of synthetic weather data is not to *predict* future weather. Incorporating stochasticity does not automatically improve the predictive power of simulation for a specific time in the future. Weather or climate predictions are imprecise and usually expected to be 'true' only in some broad statistical sense. The same is true for normative usage schedules published by standard-setting bodies such as the Swiss Society of Engineers and Architects (SIA). Rather, we expand the role of simulation in exploring design options by broadening the test conditions to explicitly calculate variation due to unknown future weather conditions. We propose to do this by simulating a building model with an ensemble of weather files, i.e. a Monte Carlo (MC) simulation. The principal difficulty, and the novelty in our work, lies in finding a time-efficient and sufficiently generalizable way to generate this dataset without requiring access to large historical datasets.

This thesis, thus, presents two contributions: (1) Synthetic weather time series that enable the explicit calculation of the uncertainty in building simulation due to weather inputs; and, (2) A customisable emulator that supplements full-scale simulation for computationally-intensive uncertainty or sensitivity analyses. The calculation of uncertainty using synthetic weather time series may be interpreted as a Monte Carlo (MC) simulation (of a building thermal model) with respect to the weather input. The influence of weather input on simulated energy consumption does of course interact with other building properties. That is to say, the work presented here is easily extensible to the examination of uncertainty due to any inputs and should thus be treated as a brief foray into a *stochastic paradigm* for BPS. Since uncertainty or sensitivity analyses are usually computationally-intensive, a strategy is proposed to construct meta-models or emulators as rapid-response supplements to full-scale simulation. In principle, the examination of uncertainty or sensitivity does not need an emulator, and using one adds error and uncertainty to the system. However, practical limits on computational time force one's hand. In addition, the technique used in this thesis to construct emulators yields an explicit distribution of the output at some unknown point, i.e., a formal quantification of the uncertainty at that point.

What does it mean then, to use simulation in a "stochastic paradigm"? The concept of a "paradigm" within which scientific enquiry is carried out comes from the work of Thomas Kuhn[3], and more broadly, philosophy and history of science. We use a loose paraphrasing of Kuhn's insights into the nature of scientific progress to explain how a paradigm shift may be interpreted. The essence of a paradigm shift is that one has to change one's perspective on a certain problem to be able to use the new techniques or procedures that are proposed as solutions. Not all of the underlying assumptions and vocabulary of the existing system will apply in the new system or paradigm. Often, a change of paradigm may directly contradict the established conventions of the previous system. In this these, the proposed shift of paradigm is more about perspective than a new theory or discovery.

If a user is looking for a precise 'answer' from a building simulation, then the techniques proposed in this thesis are useless. The 'synthetic' files proposed in this thesis are not precise predictions of the future weather. The emulators are approximate *supplements*, not *replacements*, for physics-based models. The point of an emulator is to allow the probing of a rapid-response surrogate for computationally intensive tasks like sensitivity analysis, or exploration of design alternatives with uncertainty, but it will always deliver responses with an approximate *confidence interval*. Since

[3] TS Kuhn (2012). *The structure of scientific revolutions*. 4th ed. Chicago; London: The University of Chicago Press

building simulation is an approximation of the energy used by a planned building in its lifetime, we contend that the use of Monte Carlo Analysis (MCA) with an emulator to explicitly include uncertainty improves the information available to the user. In a paradigm of deterministic simulations, for a highly non-linear system like building simulation, the user cannot know the extent to which the 'answer' from a typical file is representative of the range of 'answers' that are reasonable for a given climate.

The synthetic weather files and stochastic emulator proposed in this thesis offer a framework to: (1) quantify the uncertainty in simulation outputs; and, (2) examine the sensitivity of a building's performance to variation in various climate- and building-based inputs of interest. The concepts of uncertainty analysis are certainly not exclusive to thermal simulation, so are discussed in the wider context of BPS. The same can be said of the synthetic weather data introduced in Chapter 3: the time series may be used for modelling any system that uses weather as an input. Future research will extend the usage of these files to PhotoVoltaic (PV) systems, urban simulation, and modelling demand for electrical networks (Section 5.3).

PRÉCIS

- All inputs to building simulation are uncertain, with different levels and sources of uncertainty.
- Not all of this uncertainty can be *eliminated*, though some of it can be reduced.
- A solution/design focussed on just the mean or deterministic inputs only answers the requirements under these conditions.
- The actual conditions experienced by the building, and its as-built properties, may vary substantially from the mean.
- ■ Explicitly including the uncertainty of inputs through, for example, approximate confidence intervals, improves the robustness of design.

1.2.2 Thesis Structure

This thesis is divided into two distinct *streams*: one dealing with the creation of synthetic weather time series, and one with regression-based emulators for rapid sensitivity analyses. These two are both required to achieve the overall goal of the thesis – practical sensitivity and uncertainty analyses, especially with respect to weather inputs – but they are also usable independently.

This chapter, INTRODUCTION, discusses the motivation and background for the thesis.

Chapter 2, STATE OF THE ART summarises the extant literature about the problems of concern to this work, and solutions that have been proposed so far. It also covers the techniques and results upon which the work is built.

Chapter 3, SYNTHETIC WEATHER INPUTS FOR BUILDING SIMULATION, introduces the synthetic weather generation procedure. First, we lay out the process of extracting random and deterministic components from a weather time series, and discuss the resulting quantities. Then, we describe the process of simulating the random components to create synthetic time series. Finally, we demonstrate how these files may be used in a simulation workflow to analyse the impact of different weather conditions on the annual energy use. The chapter includes random series with and without climate-change forecasts.

Chapter 4, EMULATORS FOR UNCERTAINTY AND SENSITIVITY ANALYSES, discusses three classical (parametric) techniques, from the simplest most restrictive technique to the most generalised, and a partially parametric technique, Gaussian Process regression. The goal is to create a workable model with relatively small data sets. The techniques are discussed using two theoretical examples.

Chapter 5, CONCLUSION, summarises the contributions of this thesis and future extensions of this work. The appendices carry additional concepts and results related to each stream, Appendix A for Chapter 3, and Appendix B for Chapter 4.

1.2.3 Terms and Terminology

This thesis uses a number of concepts and techniques from different fields, so it is useful to get some terminology out of the way first. A detailed glossary/nomenclature is provided in the appendix. Longer explanations and background for some terms and concepts are also provided in the appendices (Appendices A and B).

The words 'weather' and 'climate' are not interchangeable, though they are often confused outside specialist circles. Among the many definitions available, we present a succinct statement by Essenwanger (2001): "... 'weather' is an instantaneous state of the atmosphere and 'climate' is an average state". Human reports are notoriously unreliable vis-à-vis climate reports, and when most people talk about 'climate', they are usually referring to the weather they have perceived in the last few days or, at best, the past season. Another relevant definition of climate is by the World Meteorological Organization (WMO), adopted at the World Climate Conference (1979): "climate is the synthesis of weather events over the whole of a period statistically long enough to establish its statistical ensemble properties (mean value, variation, probabilities of extreme events, etc.) and is largely independent of any instantaneous events" (ibidem).

In the context of weather data inputs for building simulation, weather 'file' implies the weather time series input. Most building simulation programs accept weather data as a separate 'file', usually a rectangular text file. Hence, the words 'series', 'file', and 'input' should be considered interchangeable in the context of weather data. Most weather data input to simulation programs is at an hourly time step, though the calculations themselves must often be run at a sub-hourly time step (due to stability conditions in finite difference networks with capacitance, for example). In this work, we deal almost exclusively with weather data at an hourly time step – giving 8760 hours in a year. We will generally reserve the word 'data' to denote measured or recorded data. Synthetic values, whether they are modelled, interpolated, or re-sampled, are referred to as 'time series' or just 'series'. This is not an important distinction, and a cause of frequent lapses, because we expect the reader to *know* that of course modelled time series are not the same as recorded *data*. 'Projections' or 'forecasts' of future climate are predictions based on computer models. The word 'projection' should not be confused with the mathematical operation of 'projecting' or mapping higher-dimensional spaces or objects onto lower dimensional basis spaces. In this thesis, we use future time series of mean values from the *CORDEX* website[4].

We will use the term 'energy need' and 'energy demand' to denote the energy requirements of a building, calculated as an ideal load. This is an integral quantity, so is usually stated over some period, like a year. It is equivalent to saying 'energy consumption', or 'effective energy'. If a mechanical system is added to the calculations, and its efficiency taken into account, then we arrive at 'final energy'. If this 'chain' of energy is continued, then we can arrive at the 'primary energy', which is the amount of energy extracted from nature. This thesis does not include calculations involving mechanical systems, so we will always deal only with energy need or consumption, stated in kWh/m^2 (Energy Use Intensity (EUI), or energy use divided by the floor area). None of these quantities should be confused with 'power', stated in Watts [W], which is not used in this thesis.

The terms stochastic and random are used interchangeably in this thesis. They mean the same thing, in the contexts that we are familiar with. Favouring simplicity over pedantry, we will try to use random or randomness, unless otherwise forced by the context. For example, 'stochastic weather files' mean time series with random com-

[4]World Climate Research Programme 2015.

ponents fed as input to building simulation. Technically, any random number generation/simulation is sampling a pseudo-random series. A pseudo-random series should approximate the uniform distribution of truly random numbers, in the interval [0,1], and the draws should be independent (Kleijnen and Groenendaal 1992).

The words façade/facade and building envelope are usually used interchangeably. For the most part, we will use the latter term in this thesis. This is to make a distinction between the facade as the veneer, or outward face of the building, in contraposition to the envelope, the layer(s) that together make up the interface between the indoors and outdoors. Facade usually only refers to the walls of a building, while the 'envelope' includes the roof, floor, and underground walls as well. In this thesis, we are interested in this concept of the envelope as a skin, so the use of facade is less appropriate.

The phrase 'Building Performance Simulation (BPS)' and the word 'simulation' should be interpreted to mean *building thermal simulation*, unless explicitly stated otherwise. According to Kleijnen and Groenendaal (ibidem), a (numerical) *simulation* is a model with a time dimension, which is the case in this thesis. The only common exception to this will be the use of the phrase 'Monte Carlo simulation', which is the sampling of some random distribution for sensitivity, what-if, or uncertainty analyses. The term *model* is used extensively in this work. Kleijnen and Groenendaal (ibidem) state that in principle, "... a model is anything that represents something else". For our purposes, this definition is far too broad. Thus we limit ourselves to the concept of "abstract models": equations, and the computer programs that solve them. A further categorisation of models is into deterministic and stochastic/random models. Deterministic models, which most building performance simulation is based on, are perfectly repeatable. In the sense that one combination of inputs will always give the same outputs, regardless of how many times a simulation is run. Stochastic or random models are, broadly speaking, not exactly repeatable because of intrinsic randomness in one variable or more. The randomness could also arise due to the use of stochastic differential equations, for example. NB: The way we use the terms determinism and deterministic should not be confused with the philosophical concept of determinism.[5]

The words 'meta-model', 'surrogate model', 'response surface', and 'emulator', all usually refer to a regression-based mathematical approximation of a real system or simulation. We will generally use only the last term, emulator, and we are of course approximating a simulation, which itself is an approximation of a complicated real system. Each of these names gives a useful insight into the purpose of these mathem-

[5] "The world is governed by (or is under the sway of) determinism if and only if, given a specified way things are at a time t, the way things go thereafter is fixed as a matter of natural law" (Hoefer 2015).

atical objects: they are models-of-models (meta-models); they are replacements or substitutes for something that is usually more cumbersome or otherwise expensive to probe (surrogates); they fit a continuous, mathematically tractable function to discrete data (response surfaces); they imitate the behaviour of some complex underlying system (emulators). Two terms often used in regression are 'parameter' and 'variable'; a variable is directly observable, a parameter is not. In addition, variables can be endogenous or exogenous: i.e., intrinsic to a model or extrinsic to it (ibidem).

Finally, a note on spelling and orthography. By force of habit, we use the British/Commonwealth variants for most spellings. For example, optimization is optimisation, color is colour, center is centre, but regression is still regression. The only exception is when we are directly quoting an author, publication, or manual. This leads to some inconsistencies but leaves us with a clear conscience. We will also avoid umlauts (Köppen → Koeppen), accents (rôle → role), cedillas (façade → facade), and other orthographic fancies, unless they alleviate confusion in a certain context.

> Content that is formatted with wider margins like this is a direct quote from another author. These quotes are intended to be part of the argument, but have not been paraphrased because the original writing succinctly conveys exactly what we were trying to say in that context.

> **NOTE ON PRÉCIS IN THE TEXT**
>
> - Whenever text is grouped like this with a bar on the left, it is either summarising the argument from a section or making a crucial didactic point, usually as part of an informal *syllogism*.
> - The ordinary bullet (•) indicates a premise or proposition.
> - The '■' denotes a conclusion.
> - Since this is not a work of formal logic, we are not particularly concerned with respecting the rules governing syllogisms. This is just a useful structure to fit arguments, and that is how it is used.
> - Despite being informal, the arguments are meant to be sound and valid, and the justification can usually be found in the text of the section where a précis appears.
>
> ■ These should be treated as informal arguments based on assertions and inferences.

1.3 Climate and Buildings

> The structure which in a given environmental setting reduces undesirable stresses, and at the same time utilizes all natural resources favorable to human comfort, may be called 'climate balanced'.
>
> *Design with climate : bioclimatic approach to architectural regionalism*, Olgyay and Olgyay (1992)

1.3.1 History, and Business As Usual

Jean Dollfus, in his sweeping review of world habitation (*Aspects de l'architecture populaire dans le monde*, 1954), concluded that dwelling types are defined less by national frontiers than by climate zones. He argued that, "allowing for some variation in local taste and tradition [like superstitions, religious beliefs, or cultural norms], the general forms of [traditional] native habitation are born of the climate" (ibidem, ref. 13-14). An analogy may be drawn between the evolution of building forms in distinct regions and the phenomenon of 'convergent evolution' seen in nature[6]. Any number of examples of pre-modern built environments from geographically separate but climatically similar regions show a remarkable convergence of urban and individual building morphology, bioclimatic strategies, and even material use. For example, the tightly clustered urban forms of the hot deserts of North Africa, *medinas*, are very similar to the tightly-packed old cities of South Asia, like Old Delhi or Ahmedabad in India[7]. The same is true of house styles in the Alps and those in the Himalayas, with their rubble masonry walls, timber structure, and slate tiles[8].

We are only claiming that climate was a *dominant* driver of the evolution of the built environment in pre-industrial societies, not the *sole* driver. For example, the twisting entrances, *jaali* screens[9], and general inward plan of South Asian courtyard houses, or *havelis*, serve to keep the occupants hidden from the view of passers-by, regardless of their climatic benefits[10]. We are also not saying that these architectural cultures developed in isolation, trade and diplomacy doubtless ensured extensive

[6] "... the process by which unrelated or distantly related organisms evolve similar body forms, coloration, organs, and adaptations" (Pianka 2008)

[7] We have avoided using images not produced by us to be able to distribute this document freely. For those not familiar with these examples, a simple search on the internet will furnish adequate images.

[8] See Footnote 7.

[9] Carved stone or wood screens, found across south and west Asia, the Mediterranean, and east Africa.

[10] Again, Footnote 7.

cross-fertilisation. Modern societies, with their efficient systems, cheap fuel, and advanced materials, are relatively free from the 'tyranny' of climate, or have at least degraded the importance of responding to the climate. However, if the designer and/or the client care enough about such things as comfort and energy use, then harnessing the climate is a sound and prudent strategy.

Précis

- Climate has a significant and complicated effect on building performance. Regardless of system sizing, climate is still the *boundary condition* that drives the system.
- Buildings are often built to last several decades, exposing them to a huge variety of weather conditions, only *some* of which can be anticipated from historical data.

■ The assessment of building performance should include an analysis of sensitivity to weather parameters, like temperature. This may be expressed as variability on the output.

The relationship between buildings and the climate is complex, and this thesis only investigates the *thermal* aspects of this relationship. Buildings have been variously described as a 'third skin', i.e., after the biological skin and clothes; as interfaces between the 'indoors' and the 'outdoors'; and other choice metaphorical objects. Simply put, the complexity arises from a constant need for *balance*. The same building (envelope) that holds in the heat well to make a cosy indoor living space in winter may become unbearably stuffy in the summer. The relatively mild winters of summer-dominated climates may be quite uncomfortable in buildings designed for shade and breeze. Heschong (1979) describes the relationship as arising from necessity because living beings, particularly humans, only survive well in a relatively narrow range of temperatures. The title of her book, *Thermal Delight in Architecture*, does indicate however that it is unusual to find a culture that stops at that. Therefore, buildings have this self-contradictory task of both protecting us from the elements and harnessing their power for our benefit, of being a barrier without cutting us off from our surroundings.

1.3.2 Climate Change

"I wish it need not have happened in my time," said Frodo.
"So do I," said Gandalf, "and so do all who live to see such times.
But that is not for them to decide.
All we have to decide is what to do with the time that is given us."

J.R.R. Tolkien,
The Lord of the Rings: The Fellowship of the Ring

Climate change introduces an additional, and very significant, risk factor in the built environment. We will by and large avoid a discussion of the necessity for *mitigation* of the environmental impact of the built environment versus *adaptation*, after this section. We take it as given that the climatic **robustness** of buildings must be improved *concurrently* with their environmental impact. It would hardly be progress to deliver high performance buildings that consume *more* energy than before[11]. Broadly, mitigation is the reduction of the impact of buildings on the environment, particularly their contribution to energy use and emissions. Adaptation is the modification of designs or operation to accommodate the slow change of mean climate, and the extreme weather events that may occur with changed frequency and intensity.

Mitigation still remains an extremely important task for the building industry, since buildings constitute up to 40% of energy use in industrialised countries (European Union 2010). Allowing for some difference in calculation methodologies, de Wilde and Coley (2012, and references 4 and 5 therein) put the contribution of buildings to anthropogenic greenhouse gas emissions at 25-40%, of which 40-95% are usually caused by operational energy use. However, adaptation to climate change is equally important for human health and safety in the built environment of the future because, regardless of how the world acts now, the widely-held view among climate scientists is that some climate change is inevitable.

The IPCC's latest Synthesis Reports (AR 5) state unequivocally that the climate is changing (IPCC 2014a,b). In these reports, possible future conditions of the global climate are represented by several different Representative Concentration Pathways. However, the report makes no claim as to which of these possible pathways or scenarios will eventually turn out to have been correct. Designing and retrofitting current buildings to meet future demands poses a tremendous challenge to the industry, since we can not know future weather precisely. Thus, we limit the discussion of climate change in

[11] Though that was exactly the definition used in the first half of the 20th century!

this thesis to an exploration of its contribution to uncertainty in building simulation. In the case of weather inputs, the state of the art calls for a simulation of thermal performance based on typical weather data, a Typical Meteorological Year (TMY) or Design Reference Year (DRY), although there has been some research questioning their utility, e.g. Crawley (2011), Crawley and Huang (1997) and Crawley and Lawrie (2015b).

The difficulty of accurately predicting the behaviour of a system that depends on the climate with normative (historical) data is that we cannot fully characterise the climate itself, especially future climate. In other words, while the distribution of a weather parameter's past values (e.g., average and extreme daytime temperatures) can be known from historical data, using past data without any modification represents an assumption about the future: that of a stable climate.

PRÉCIS

- Climate change will affect the built environment across the world, with effects ranging from mild warming indistinguishable from the effect of urban heat islands, to severe heat waves of previously unknown intensity, and/or increased frequency.
- Buildings consume a significant share of global energy production, and this figure will only grow (in absolute terms, not necessarily as a portion of the total usage) as the rest of the world catches up with the developed world.
- Building design must consider both adaptation to, and mitigation of, climate change.

1.4 Energy-Conscious Building Design in an Uncertain Climate

Buildings don't use energy: people do.

Janda (2011)

Energy-conscious building design, in the context of this thesis, is design for *efficient operational energy use for space heating and cooling.* We are not looking at embedded or grey energy, during construction or demolition. Nor are we concerned with the so-called 'plug loads', i.e., electrical energy use for appliances, lighting, and processes, except for the internal heat gain from these. It is also assumed that the building is designed to be comfortable, appropriately ventilated and lit, preferably with daylight to create desirable interiors and save even more energy. This throws up two questions which we will address in this section: what is ‘comfortable’, and how do we know if a design is going to be comfortable (appropriate for our indoor environmental goals)?

1.4.1 Comfort and Expectations

It is important to acknowledge that energy-conscious design only exists in the context of comfort expectations and the means to achieve them. That is to say that energy usage for space heating and cooling is only required if certain thermal conditions are desired, and these cannot be achieved through the design alone. From this perspective, the most efficient building is one that uses *no* mechanical energy for heating, cooling, and ventilation. Hence, we begin a discussion about energy-conscious building design with a brief note on thermal comfort. Once that is out of the way, we will proceed to ignore it for the rest of the thesis, assuming that the user will adjust all methods to account for different static or adaptive standards of thermal comfort (see Section 2.3.1). Future work will explore the interaction of comfort and expectations with weather uncertainty, and the performance of regression models with these ‘extra’ factors (Section 5.3.3).

Thermal comfort may be defined as “that condition of mind that expresses satisfaction with the thermal environment” (*ASHRAE Standard 55-2010*). This statement is somewhat vague because it is an acknowledgement of the fact that comfort arises from a complex interplay of physical, physiological, psychological, and social factors (Grondzik, Kwok et al. 2011). It would not be a stretch to say that much of what determines an individual’s comfort with their surroundings is down to *expectation*. Expectation is not amenable to measurement, though, so thermal comfort research has focussed largely on heat transfer/exchange models of the human body and its surroundings, so far. In this thesis we use the static American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) comfort model, using a pair of set-points for heating and cooling in each simulation. That is, the “target” temperature is fixed, regardless of outside conditions or time of year. The settings for each case study are not particularly important to the work presented in this thesis, and interested readers will find the original Energy Plus input files (IDF) with the archival entry of this thesis (infoscience.epfl.ch).

Précis

- We assume that a building is designed to consume as little energy as possible.
- If people do not demand certain indoor conditions, i.e., they accept the indoor conditions created passively by the building, then there is no need to use energy.
- People *do* demand that indoor conditions stay in a certain band, which may be fixed at some arbitrary value for a particular design exercise, based on one's choice of comfort model.

■ Energy use can only be *minimised* under some arbitrary assumptions on comfort.

1.4.2 The Role of Building Performance Simulation

> Achieving a high quality indoor environment at acceptable cost has always presented a challenge for the construction industry. With aspects of sustainable development now being added ... this challenge is set to become even more formidable. Simulation represents a possible solution ... by enabling comprehensive and integrated appraisals of design options under realistic operating conditions.
>
> *Energy Simulation in Building Design*, Clarke (2001)

The design of any building involves a complex array of decisions with cascading effects. Are the energy savings from using triple-pane windows instead of double-pane ones enough to cover the costs of installation? Or, would increasing the insulation level substantially decrease the need for heating in the winter without causing problematic overheating in the summer? Or, how should the building be oriented to maximise the view outside enabled by the relatively well insulated triple-pane windows and walls? However, merely using complicated software is not the end of the story. The ultimate prospect of Building Performance Simulation (BPS), is that "of a truly powerful computational approach to design whereby arbitrarily complex models may be evolved on a task-sharing basis, such models readily exchanged and understood by others, industry standard assessments automatically invoked, and seamless integration within the temporally evolving design process assured" (Clarke 2015). At the moment, however, it is probably more accurate to say that the use of simulation is limited to "... code compliance checking or thermal load calculations for sizing of heating, ventilation and air-conditioning systems in detailed design" (Hopfe 2009).

PRÉCIS

- To minimise the energy use of a building, the planner has to be able to predict the impact of design decisions under the given site and usage conditions.
- It is time consuming, expensive, and extremely impractical to keep experimenting with systems and components once a building is built.
- Simulation is an efficient means of conducting what-if analyses like these. In the case of building simulation, we are almost always interested in future conditions and responses.
- Building simulation is useful for energy conscious design because it enables "the emulation of future realities at the design stage" (Clarke 2001).

1.5 Quantifying Uncertainty, Sensitivity, or Both?

Ultimately, building simulation is just that: an *imitation or simulation of reality*. Improving the fidelity of the model, i.e., the representation of a design and site in a software, improves estimates of the likely (energy) impacts of decisions. However, simulation is neither a perfect prediction, nor a substitute for judgement. It may provide *guidance*, but the work presented in this thesis does not take this literally. Instead, we take the position that it is good practice to account for the possible effects of the uncertainty about, and the variation of, the values of inputs. Specifically, in the domain of this thesis, these inputs are the weather parameters, and their interaction with the building envelope. Throughout this thesis we will discuss methods to conduct uncertainty and sensitivity analyses. These two concept are related in this context, but they are not the same.

1.5.1 Does Simulation Imply Prediction?

Seasoned users of simulation will argue that prediction is not the point of BPS at all. Rather, it is the exploration of design alternatives through a series of what-if scenarios. While we agree with this in principle, it is unusual to find a simulation exercise *not* being interpreted as a prediction. Interpreting the output from a simulation run as the prediction for a specific point of time in the future is certainly incorrect. Two particular boundary conditions – weather and occupancy – will almost certainly not be the same at a specific hour in the future as they are at the same hour in a simulation run based on typical data and normative schedules. Yet, we expect that the output from a simulation using typical or normative data is a valid *mean* response, i.e., the response of a building to typical or mean conditions[12]. A user also expects that the differences between the energy use of alternative designs obtained from simulation with typical inputs are similarly indicative of the mean/median difference. This too is usually the case. However, the mean difference is *not* representative of the *spread* of the difference between two options. A mean or median is always to be interpreted as a single-point estimate of a sample distribution – its representativeness is a subjective matter. This also implies, however, that it is entirely conceivable for the difference between two options to be statistically *insignificant*.

In a review of literature on predicting building energy consumption, Zhao and Magoulès (2012) point out that the complexity of predicting energy use makes precise prediction difficult. In this thesis we take the position that, while *precision* is a matter of choice, *accurate* prediction of future energy consumption beyond very short time horizons (usually within a day) is impossible or, at best, impracticable. One can say that a building will consume exactly ~38 kWh/m^2 of energy for heating next year (Figure 1.1), a very precise prediction based on a typical weather file[13]. But is it *useful*, when simulations of the same model with recorded weather data over 20-odd years show variations between ~30-48 kWh/m^2 (Figure 1.1 again)? We argue that prediction must strike a balance between precision and accuracy as they seem to follow a relationship not unlike that of momentum and position in Heisenberg's uncertainty principle. The more precise a future value of energy used is, the less likely it is to be accurate or correct. The more likely to be correct or accurate one seeks to make a prediction, the less precise it inevitably ends up becoming[14].

When dealing with inputs that are not known with absolute certainty, it is almost tautological to say that the precision and accuracy of outputs depends on the precision and accuracy of inputs. In the context of building simulation, even if the representation of a building in a software is completely reconciled to the design, or as-built conditions, neither future weather conditions nor occupancy and usage are exactly predictable. Weather or climate predictions cannot be exact, based on the current state of knowledge, and are usually expected to be "true" only in some broad statistical sense. In this thesis, we make the case that this is how the outputs from the modified building simulation workflow introduced in this thesis should be interpreted.

[12] Not *always* though, because it is ultimately influenced by the quality or representativeness of the typical inputs.

[13] Details of this energy model are in Table B.4, and the codes on the x-axis correspond to the last two digits of the entries in the second column of that table (new codes). This simulation study is referred to as Case 1 in Chapter 4.

[14] We will occasionally draw analogies with quantum mechanics when talking about probabilities of future weather. However, these are merely intended to be figurative, since quantum effects are nonsensical at anything larger than the sub-atomic scale or smaller than a stellar scale, which is the comfortable in-between where we study the built environment.

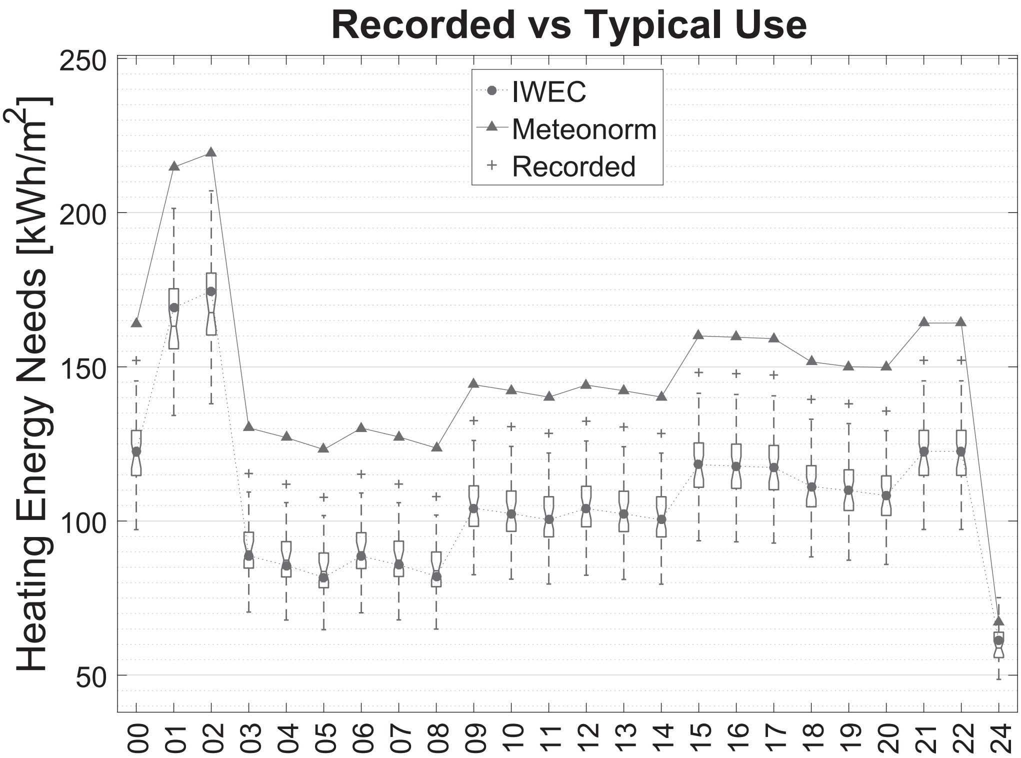


Figure 1.1 – *Renovations for a single-family home, options 01-24, simulated with recorded weather data from Geneva. Case 00 is the base case, or the house as it is now. Each number on the x-axis represents a possible refurbishment option, as detailed in Table B.4. The circle and triangle represent values from simulations with typical weather files, while the box plots are constructed from simulations with recorded weather data from 1988-2013. This graph is meant to show the variations in (calculated) building energy usage due to weather, especially when juxtaposed with variation due to refurbishment interventions.*

In this thesis we show examples which give credence to our hypothesis that, depending on the building in question, the spread of possible energy outcomes is so large as to make the difference of means between two design choices misleading. A precise answer (the difference of means) is inaccurate for the majority of possible outcomes (the difference of spreads). The difference shown by simulation with a single typical file often becomes irrelevant when the entire sample, based on several weather files, is taken into account[15]. For example, the difference in mean values between the base case and first renovation in Figure 1.1 is ~10 kWh/m^2 (48-38), whereas the differences caused by using two different typical weather files is already ~14 kWh/m^2 (52-38). The

[15] Formally speaking, the variation around a group mean may be large relative to the variation of group means around the overall mean (The MathWorks, Inc. 2015).

range of consumption values over the recorded data is ~28 kWh/m^2, for a period of about 20 years.

1.5.2 Why Abandon Certainty?

... All hope abandon, ye who enter here.

Dante Alighieri (1265–1321),
The Divine Comedy (Inferno, Canto III)Harvard Classics, translated by Henry F. Cary

It has been previously proposed by us and colleagues (Chinazzo 2014; Chinazzo, Rastogi et al. 2015a,b) that a 'range' of possible performance outcomes, i.e., the results from simulation runs with different weather inputs, is a better characterisation of the *range* of performance that a building will inevitably give. Simply summarised, the argument is that if one does not know exactly what (weather) inputs one's (building) system will experience, then one is better off knowing the effect of a range of possible inputs. A given weather file is, after all, a representation of one scenario out of an immense number of possibilities. Therefore, by using only one weather file, we are restricting ourselves unnecessarily to one "experimental result". If a building is never going to experience a narrow set of weather conditions exactly, e.g. the ones contained in a typical weather file, then the quality or 'averageness' (or ability to represent best the most typical weather) of said weather file is irrelevant.

Long simulation with measured data does seem, intuitively, to be a better depiction of climate conditions than a single "typical" time series. However, it does not guarantee coverage of future conditions. In other words, while the distribution of a weather parameter's values can be known from historical data, using historical data represents an assumption about the future: that of a stable climate. On the other hand, despite the unequivocal conclusion from the IPCC's latest Synthesis Report (IPCC 2014a) that the climate is changing, it is not knowable which of the possible scenarios they list will eventually pan out. While a long record does enable a Sensitivity Analysis (SA), one is still hostage to the vagaries of the weather when using it. That is to say that there are several possible future conditions that may not have occurred in the recent past. There are no guarantees about what conditions may prevail in the future based on knowledge of past conditions. As far as we are aware, the temperatures of future years do not have to follow some well-defined mathematical relationship with temperatures from previous years, or even some well-defined periodic relation.

Assuming one has access to long-term hourly data from a weather station that is sufficiently close to the area of interest, in addition to a typical year file, one can know how a building *would have* behaved. However, one has no tools for assessing any arbitrary weather conditions. One might have a reasonable idea of the expected range of average temperature rise in a climatic region, thanks to the IPCC's publicly available models, but one does not know the possible implications of this at an hourly resolution for a given weather station. Our approach seeks to address this incertitude by proposing a 'what-if' analysis of a building to variations in the climate. In other words, we introduce stochastically-generated synthetic weather data without seeking to forecast the 'true' future climate. Analysing the sensitivity of buildings has important ramifications for planning, policy, and risk assessment (more on this in Sections 2.4 and 2.7). For example, it could be useful to know the distribution of total energy in some applications, or the 95% confidence interval of the mean energy use. In other applications, the extreme values and their estimated probability might be more useful. For example, in studying overheating risk or the risk of exceeding some level of peak demand.

PRÉCIS

- A building will almost certainly not experience the exact sequence of weather in a typical file.
- Historical data, when available in a convenient form, only *partially* characterises future weather conditions.
- Deterministic (exact) models of weather are useless beyond very short time horizons.
- Working with an ensemble of plausible weather years is a better characterisation of the (climatic) boundary conditions that will be seen in a building's lifetime.

1.5.3 Numerical vs Analytical Approaches

If building simulation algorithms could be satisfactorily characterised by a simple emulation, like a polynomial, then the calculation of sensitivity and the propagation of uncertainty would be trivial *plug and chug* affairs. Similarly, analytical confidence intervals could be used to quantify uncertainty if one were sure of the limiting distribution of an output or input. Since neither of these is the case, we take a numerical (read: computationally intensive) approach. The use of regression-based emulators serves both to reduce the computational load of sensitivity analyses, and, in the case of Gaussian Process regression, to chalk out explicit confidence intervals.

Numerical approaches themselves introduce uncertainties and errors (more on the difference between the two in Section 2.4). On a conceptual level, using an emulator means the user is now measuring the sensitivity of the emulator or the propagation of uncertainties through the emulator, and not the original computational system.

1.5.4 Uncertainty Analysis

Often called uncertainty *quantification* (UQ), this provides a framework to construct "computational error bars" (Iaccarino 2008). It is useful for decision-making because it quantifies what the *user does not know*, and in many cases *can not know*. When we query an emulator constructed with the Gaussian Process regression technique proposed in this thesis, we obtain an estimate of the mean and variance of a Normally-distributed output variable, e.g., cooling energy, at that combination of inputs. The confidence interval in this case is exact, conditional on the validity of the fit (surrogate). If this fitting step is not used, then a user could collate the results of a set of simulations based on an ensemble of synthetic weather files. The confidence intervals on the output constructed in this manner would be *approximate* or *empirical*. In practical terms, there will not be a large difference between the two kinds of confidence intervals, though the justifications are quite different. This interval would be representative of the confidence in the output based on an uncertain input – regression inputs or predictors in the former and just the climate in the latter case.

Simulating every combination of weather file and building properties of interest could also be used to estimate of the risk of some condition occurring, say overheating. Once again, the output from a Gaussian Process (GP) emulator would be an *exact* confidence interval, say the range of temperatures which could occur with 95% confidence. A collation of the simulation results would give an empirical confidence interval with approximately the same result. Thus, in our case, uncertainty quantification is a method to overcome the fact that the weather inputs are inherently vague, or that a fair amount of information has been lost in the creation of 'typical' weather data. The work summarised in this thesis is explicitly quantifying the confidence in the output due to a particular combination of inputs, i.e., building properties and aggregate climate conditions, by constructing either exact intervals on an emulator, or approximate (empirical) confidence intervals from the original simulation.

PRÉCIS

- The system being simulated here – a building, its site, and its users – is incredibly complex and analytically intractable.
- Many of the inputs, if not most, are uncertain and/or pseudo-random.
- Mean values can not be used to create confidence intervals because the true distribution of the outputs is unknown.

■ *Numerical* uncertainty analyses should be included in the simulation workflow.

1.5.5 Sensitivity Analysis

These kinds of analyses inform the user about the significance, or lack thereof, of inputs. One must be conscious of the uncertainty in the inputs and the system, but a sensitivity analysis of an output to some inputs can proceed just fine without it. Say one simulates a building[16] with the synthetic weather files proposed in this thesis. Examining the variation of the output, e.g., cooling energy, against variation in some appropriate weather-based input, e.g., Cooling Degree Day (CDD), quantifies how sensitive the building is to temperature, represented by the aggregate metric Cooling Degree Day (CDD). Using sensitivity analysis may help the user focus on important parameters that should be addressed, because the user will know what impact the variation of a certain input has on the output.

Using the synthetic weather inputs proposed in this thesis provides an estimate of the *distribution* of some output, like energy, conditional on the range of these weather input values. Based on the frequentist interpretation used in this thesis, approximate confidence intervals may be drawn around an output through simulation of the building model with a set of related but varying weather inputs. Hence, a sensitivity analysis also lets us assess the uncertainty of a certain output, based on the perturbation or variation of input(s). In this case the Gaussian Process regression enables a rapid estimation of the outputs, i.e., the range of energy use values possible when the weather parameters are in some ranges of interest. It does not add anything to the sensitivity analysis conceptually – it only lets the analysis be performed rapidly.

[16] Recall that we are talking about building performance simulation.

PRÉCIS

- The system being simulated – a building, its site, and its users – is incredibly complex, with many unknown and/or pseudo-random inputs.
- In a complex system, the relationship between inputs and outputs is often analytically intractable. That is, there is no simple way of knowing what effect an input will have on the output.
- Yet, these inputs may affect the output in significant ways. Using only mean inputs does not permit an analysis of these effects.

■ *Numerical* sensitivity analyses should be included in the simulation workflow.

1.6 Simulation Inputs: Perspective and Interpretation

... there are known knowns... things we know we know.
We also know there are known unknowns ...
we know there are some things we do not know.
But there are also unknown unknowns – the ones we don't know we don't know.

Donald Rumsfeld
United States Secretary of Defence. February 12, 2002

1.6.1 Weather Inputs

Every physically-reasonable value of a weather parameter is *possible* in a given location, with some finite probability. However, the *probability* of getting a specific value in the immediate future, say the next century, is the tricky bit. It provides part of the justification of this thesis, and sustains the funding streams of entire departments. Given this uncertainty, both recorded and synthetic weather should be treated as a possible set of conditions that a building may experience. Has a temperature greater than 40°C ever been recorded in Geneva? No. How long have reliable temperature records existed, for anywhere? Since the beginning of the 19th century, more or less[17]. With a changing climate and increasing length of record, are we likely to record extremes and episodes that haven't been seen in the past century or so? Yes, almost certainly. This makes separating a ridiculous value from one that is merely

[17] The oldest temperature series we found are the Central England temperatures, which go back to 1659 (Manley 1974). The oldest climate 'records' are probably the Nile flooding records, although these are patchy (Bell 1970).

uncommon, is an extremely difficult judgement call. Paleo-climatology offers an interesting avenue for future work, since it opens up the possibility of comparing climate over time scales that are longer than the geological blink of an eye that is the Industrial Revolution. Another extension to our work is the use statistical techniques to deliberately generate extreme values, even ones that would be currently considered ridiculous. Existing work of this sort on the risk of flooding and other disasters offers paths for collaboration.

We extract the essential characteristics of a climate (e.g., autocorrelation, means) and build any number of synthetic files by modelling structures in the apparently random components of the time series. This is possible because of the idea, developed by several authors (summarised in Section 2.6.3), that weather time series can be decomposed into characteristic seasonal components and apparent 'innovations', though without any claim to know the source of these random changes or innovations. This method borrows heavily from Boland (1984, 1995), Hansen and Driscoll (1977) and Magnano, Boland et al. (2008). The major difference between the models proposed by these authors and our model is twofold: we restrict ourselves to using Typical Meteorological Year (TMY) or Design Reference Year (DRY) files, while they used recorded data; and, we aim to demonstrate a generally applicable model for creating synthetic weather data for simulation, whereas they were only working with a sample of stations in a particular region. Other differences between our method and theirs are mentioned in the text where relevant. Much of the work is also based on the discussions and examples presented in Davison (2013) and Davison and Hinkley (1997).

PRÉCIS

- Future weather cannot be predicted exactly, using current methods.
- A what-if analysis, based on an ensemble of possible future weather files, will inform the designer of the consequences of certain decisions on the risk of some parameter (e.g., indoor temperature) reaching a certain value (e.g., 35°C).
- The plain synthetic files proposed in this thesis provide a set of time-unspecific variations on a stable climate.
- The future synthetic files use low-frequency forecasts of projected daily means in conjunction with noise series based on historical data to create time-specific future weather scenarios.

■ The synthetic weather files help to determine the probable severity of indoor conditions, and the likelihood of having them in this century (i.e., up to 2100).

It is tempting to fit limiting distributions on physical values, and many authors do (see Sections 2.4 and 2.6.3), but we almost completely avoid it. The trouble with choosing a theoretical distribution is that one imposes a model on the data that may or may not be correct. Instead, the non-parametric approach described in this thesis tries to create variations on the small dataset it is given. This leaves less room for generating extremes, and is somewhat hostage to the quality of the generating data. However, we found that the approach is robust enough for our purposes, and generates several episodes of interest.

1.6.2 Building Properties

Although this thesis focusses on the weather input to simulation, it is easy to extend the approach to other inputs. Building properties are conventionally regarded as fixed. That is, one assumes that the value of a (thermal) property is known because it has been quantified through experimentation. We argue, however, that materials properties should also be regarded as samples/estimates from a population input. For example, the unit conductance [k W/m^{3}K] of a piece of extruded polystyrene may be regarded as known perfectly (if experimental error is ignored). However, the conductance (U-value) of the actual insulation layer [W/K] in one's building should be understood as a single estimate of a true value. If enough buildings with the same walls are modelled, the conductances of the wall insulation in all of those buildings will be slightly different due to specification and installation errors, and over time due to degradation. This means that the actually experienced value of insulation in these buildings varies from that calculated by multiplying the unit conductance with the volume of insulation used (even if we assume that the volume is known perfectly, which it is not). Therefore it is reasonable, in our opinion, to consider the calculated value as an estimate of the true population of wall insulation conductances that will be seen by a building. This thesis does not go into how the details of interpreting a property (Bayesian vs frequentist), limiting ourselves to a conventional sensitivity analysis demonstrated in Chapter 4. More details about interpretation, and previous work in this area, are in Chapter 2.

An obvious example of a 'building' property that can be thought of as random is user input. User input is, in this thesis, aggregated into just one quantity: sum of internal heat gain (*SumIHG*). This is the sum of heat contributed by building occupants to the heat balance. This comes both from metabolic heat gains and from interactions with appliances and lights. Internal Heat Gain is conventionally calculated from a fixed occupancy profile which is based on the usage of a building, e.g., office or home.

The quantity *SumIHG* is classified as a building property but it is also an estimate of a population parameter, say $SumIHG^*$, the sum of heat gain due to any profile chosen from a population of possible occupancy profiles. Since the 'true' occupancy profile of a building is *un-knowable*, except maybe in a forensic study, it is prudent to establish the effect of an ensemble of likely usage profiles on the energy usage of the building.

2 State of the Art

If I have seen further it is by standing on the sholders of Giants. [sic]

Isaac Newton (February 15, 1676),
in a letter to Robert Hooke

References to publications on which our work is built, or which seek to answer closely related questions are scattered throughout the thesis. In this chapter we will devote most of the text to summarising these publications. Other work that has informed the general development of the thesis, including textbooks, are summarised in lists or short précis, without reference to specific conclusions or proposals in their content. Apologies are due to the authors whose work has been shoehorned thus. Well-known mathematical concepts, like regression or time series models, are in the appendices (Appendices A and B). The list of work familiar to us is limited, unfortunately, to English-language publications.

Informal knowledge transfer between masters and apprentices has been at the core of vernacular building traditions for millennia. New techniques and designs were created through a painstaking process of trial-and-error, as lessons learnt by one generation of master builders and craftsmen were passed on to another through apprenticeships. With the formalisation and specialisation of various building-related professions, publications by professional bodies and experienced practitioners now provide more pointed guidance for their respective trades. Several catalogues and guides of ‘best-practice’ for indoor environmental design have been published throughout the world.

2.1 Climate Classification and Characterisation

The air changes every day.
Every day, clouds move around, rain comes and goes, and winds change.
And every day, people all over the world try to figure out
what the air is doing and where the rain will go next.

Randall Munroe, *Thing Explainer: Complicated Stuff in Simple Words*

As we mentioned in the discussion on terminology in Section 1.2.3, climate is different from weather. Broadly speaking, climate is the collection of descriptive statistical properties that emerge from long term weather records. Weather, on the other hand, is the instantaneous state of the atmosphere.

2.1.1 Historical Development

> Classification of the climate is a grouping of atmospheric conditions for locations which show similar climatic conditions (climate types) separated by defined boundaries applied to one or more meteorological elements.

General Climatology 1C, Essenwanger (2001)

Like many scientific fields where measured quantities intersect with human experience, the study of climate and its effect on human beings has ranged from the insightful to the wacky. The earliest known efforts to divide the earth into climatic zones are from the sixth century BCE. The word 'klimata' appears in Greek texts around 500 BCE, though it refers only to length of day (Essenwanger 2001; Ward 1905). In his *Meteorology* (Book II, Chapter 5, c. 350 BCE), Aristotle describes five climatic zones based solely on latitude (climes). The zone between the tropics of Cancer and Capricorn was the Equatorial 'Torrid' zone, while the Arctic and (hypothetical) Antarctic Circles were the 'Frigid' zones. Both were considered uninhabitable. The Northern Temperate zone was taken to encompass most of the known world at the time, i.e., Europe and Asia (excluding the Far East), and the only one deemed fit for human habitation. Aristotle also observes, quite presciently in our opinion, that "... the greatest heat [in the Northern Temperate zone] is developed not when the sun is nearest to the north [summer solstice], but when its heat has been felt for a considerable period and it has not yet receded far". The presence and habitability of a Southern Temperate zone (the Antipodes) was a matter of debate until the 17th century. The geographer Ptolemy

(AD 100-170) used seven climes (latitudes) for his division of the Northern hemisphere based on the length of the longest day(Oliver 2005; Ward 1905). In fact, the literature is awash with speculative or outdated references to ‘habitable’ and ‘uninhabitable’ zones, isotherms, climate divisions, and sundry climate-specific recommendations, including Sacrobosco’s *De Sphaera Mundi*, the *Shilpa-Shastras* (Acharya 1928), etc.

These were what we would call ‘macro’ classification systems and, in the reckoning of the authors, global. The concept of ‘micro-climates’ shows up around the 13th century AD, with Magnus’ distinction of coastal sub-areas and north and south sides of a mountain, to supplement the seven-zone Ptolemaic system that lasted until 1500 AD (Essenwanger 2001). Essenwanger (2001) and Oliver (2005) trace the development of climate zones from these early works, which were almost all based on the length of the day, to modern systems based on temperature and rainfall values, and their duration, through the work of Koeppen, Humboldt, Dove, Herbertson, Miller, and others. It is interesting to see that the so-called “classical age” and “ Koeppen era” of classification coincide with the broad scientific trend, from the Enlightenment onward, to want to catalogue nature. Newer instruments and wider data-gathering propelled ever-more reworking and remixing of older classification systems with bigger, more multi-variable, formulations and datasets.

Despite being a “100 years old, the classification of climate originally formulated by Wladimir Koeppen and modified by his collaborators and successors, is still in widespread use... by researchers across a range of disciplines as a basis for climatic regionalisation of variables and for assessing the output of global climate models” (Kottek, Grieser et al. 2006; Peel, Finlayson et al. 2007). In its broadest sense, this classification scheme uses natural vegetation as a benchmark of climate type. At first glance, this method does not seem to be applicable to building design. However, natural vegetation is usually representative of several climatic parameters important to buildings, namely: temperature, precipitation, humidity, solar availability, and their seasonality. Natural vegetation is also influenced by the surface energy balance of a location (see Glossary). The surface energy balance is strongly correlated with the energy flux potentials experienced by a building. Alkhalaf and Kraus (1993) found that the surface energy balance characteristics of distinct Koeppen-Geiger climatic regions are unique, regardless of location. This is an interesting result for thermal analyses of buildings, since the surface energy balance is far more important for thermal loads than just the natural vegetation of a location.

Miller (1961) says that while vegetation is broadly representative of climatic regions, it is only the first step towards what should be a more refined system. Other factors

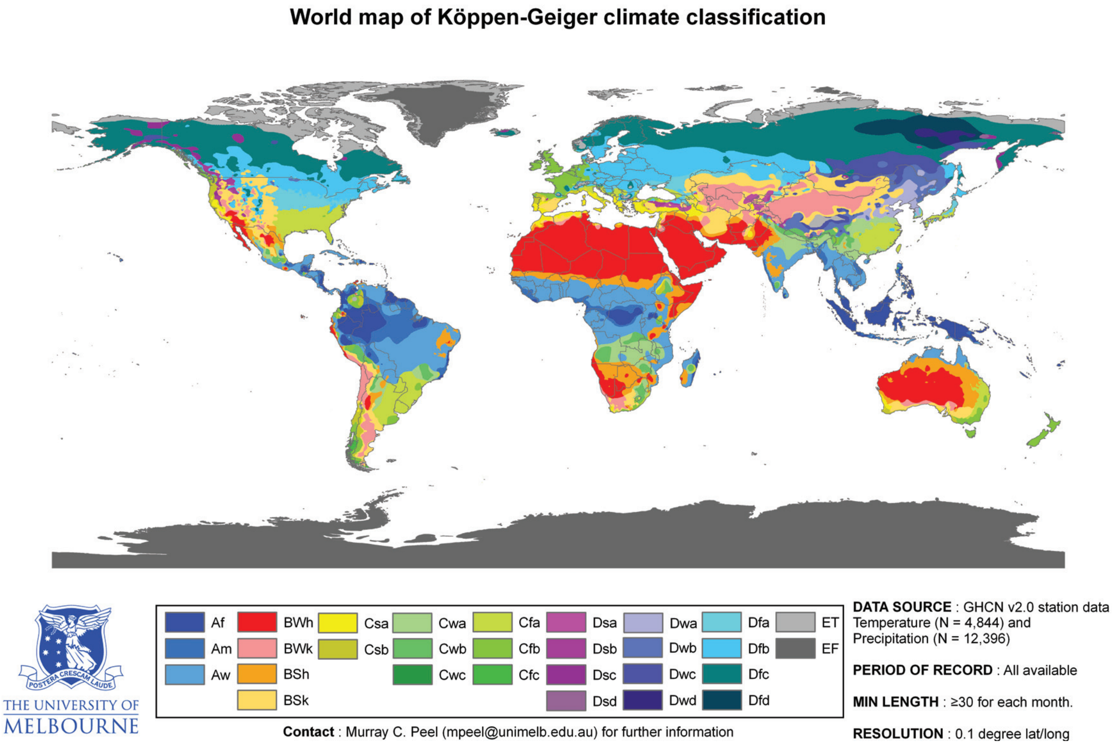


Figure 2.1 – *An updated map of Koeppen-Geiger zones from Peel, Finlayson et al. (2007).*

such as soil conditions and groundwater supply influence the vegetation in addition to climate, and these are not important to buildings. They review attempts by others (including Supan, Koeppen, and others) to create climatic division systems based on temperature, vegetation, etc., and propose a climatic division system of their own. It consists of seven general types of climate, some of which are further subdivided into two or more subtypes. Coming after the end of the classical and Koeppen eras, their conclusions have a (probably unintentional) prescience vis-à-vis building design: that *any attempts to define climatic regions based on a single climatic element are seldom satisfactory, and that they should be used carefully and interpreted liberally.*

Olgyay and Olgyay (1992) discuss some early development of climate classification and literature in the search for the 'ideal' conditions for human flourishing. There seems to have been general agreement up to the modern era that the tropics and poles are uninhabitable, or at best, stultifying. This is clearly not the case for the former, as people have been living in the tropics since the dawn of humanity, and contemporary tropical societies have super-charged, air-conditioned economies. From the Ancients onward, there has been a persistent trend to characterise one's home or preferred climate as offering the best conditions for humans to thrive, not just physically but apparently even morally (for example, Ellsworth Huntington's *Principles of Human Geography*). Modern climate classifications can be called aggregative, in that they deal with annual statistics such as the average temperature of the coldest month or the amount of rainfall received. We discuss a few systems that are relevant to building design or were derived specifically for it. These systems are distinguished chiefly by their intended application and the factors they consider.

> The almost unlimited combinations of climatic factors acting on an almost infinite variety of topography produce a bewildering number of geographical climates, and it is clear that any system of classification adopted can recognise only the broadest types unless it is to become unwieldy. But in spite of the seeming complexity it becomes clear on closer examination that certain combinations of climatic elements repeat themselves with some degree of regularity in different parts of the world, and it is convenient to recognise each type and to give it a name.

Climatology, Miller (1961)

2.1.2 Contemporary Work

2.1.2.1 ASHRAE Climate Zones

By far the most important classification system for our purposes is the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) climate zone system. ASHRAE assigns a single label to weather records from any station in the world based on temperature (as expressed through Degree Day) and precipitation. The Degree Days are used to assign the primary numerical identifier to a location (numbers 1-8). This identifier is further refined by criteria based on precipitation, giving subcategories A (moist), B (dry), and C (marine). The calculations that form the basis of the ASHRAE system are open for public view and comment, and the organisation itself publishes lists of cities in various zones in some of its standards. The full set of climates explored in our work include at least one city from each ASHRAE zone. The three primary examples presented are Geneva (4B), Chicago (5A), and Delhi (1B). (ASHRAE, ANSI and IESNA 2010; Owen 2009, 2013)

ASHRAE climate zones are widely disseminated in the building community through published design guidelines for building components and systems, e.g., *ASHRAE Advanced Energy Design Guide for Small Office Buildings*; *ASHRAE Advanced Energy Design Guide for Small to Medium Office Buildings*, and standards, e.g., *ASHRAE*

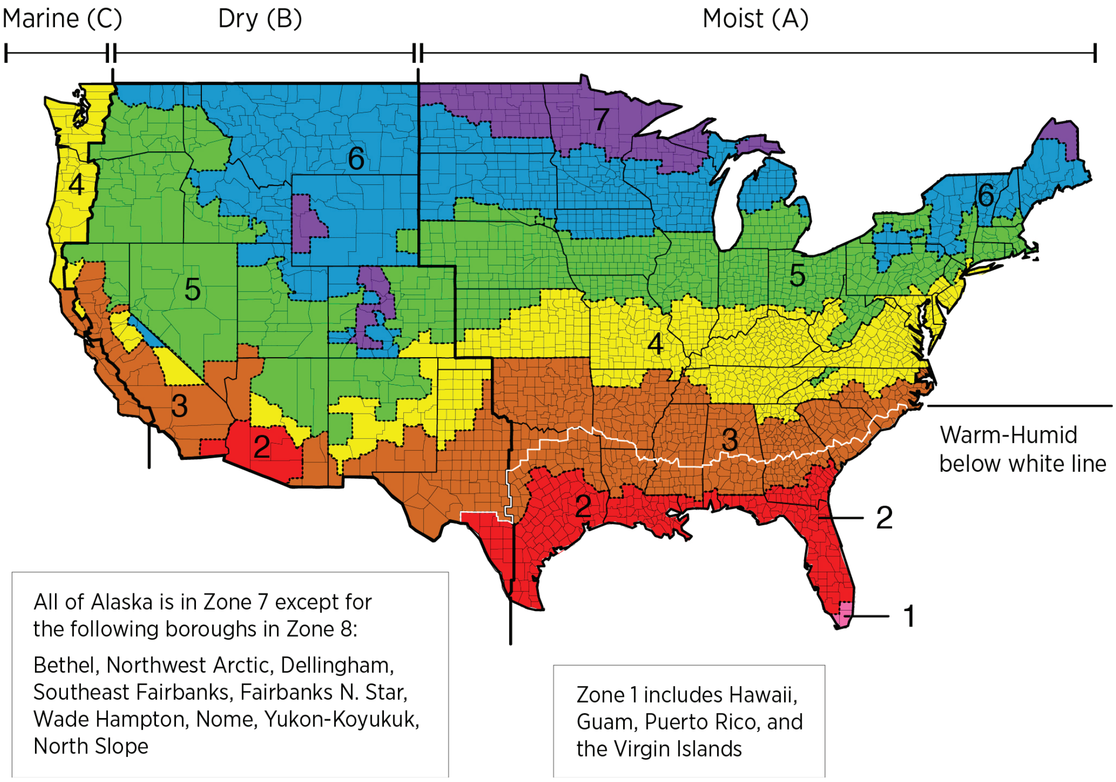


Figure 2.2 – *ASHRAE zones in the continental USA prepared for the* International Energy Conservation Code*, published by Pacific Northwest National Laboratory and Oak Ridge National Laboratory (2010).*

Standard 140-2011; *ASHRAE Standard 90.1-2010*. It is also used in the *International Energy Conservation Code* to demarcate climate-appropriate minimum requirements.

2.1.2.2 Climate Severity Index

Clarke (2001), Markus, Clarke et al. (1984) and Markus (1982) proposed a system based on simulating the reactions of buildings to climate conditions: a "Climate Severity Index (CSI)". The aim was to inform public policy on fuel subsidies and codes, help make rational decisions on capital allocations for retrofits, and aid the selection of housing sites. The index tries to synthesise the "stress placed upon a building's energy systems... by any given environment". The authors rightly point out that a possibly infinite combination of weather parameters would create the same reaction in a building, i.e., the same energy demand. And that CSIs already exist, e.g., the indices of thermal comfort, or various techniques based on Degree Days.

The index is based on regression relationships created from simulation experiments on common/standard house constructions. The climate parameters included are air temperature, 'useful' radiation, and wind. The building characteristics considered are "mass/insulation, solar 'admittance', and wind permeability characteristics of houses". Both groups of factors are very similar to the building and climate properties considered in this thesis. The regression relationships can be used to determine the effect of a very large number of climate conditions on several types of houses (27), varying in their construction. The authors construct the CSI in a 4-stage process: (1) Establish the maximum and minimum loads due to each climate variable on each building by inputting the respective extreme climate conditions in the regression equations. (2) Compute the ratio of these individual differences to the overall difference, for each house type. (3) Rescale the individual ratios from 1 (best climate conditions) to 10 (worst). (4) Produce a "family of curves... relating climatic severity to house energy requirement for the 27 selected house types".

The ideas underpinning the development of this index are similar to what we propose in this thesis. The difference is in the intended application. Instead of making a set of general regression relationships, our work focusses on providing the tools to create emulators for each design case. If the application was the assessment of climatic severity at a regional level, then Gaussian Process regression could still be used to create an ensemble of regression relationships focussed on common housing types. Together these would let the user explore the interaction of the local climate with particular construction practices or retrofit measures to make informed decisions about the portion of the building stock to be targeted for renovation or fuel subsidies.

2.1.2.3 Other Classification Systems

Olgyay and Olgyay (1992) present a simplified version of the Koeppen system with four zones: hot and humid, hot and arid, temperate, and cool. They base their simplification on the assumptions that the distribution of natural vegetation, while very representative of the climate, is not directly applicable to housing. In their view, a simplified system which considers only temperature, solar availability, and humidity/precipitation should suffice. This is a prescient observation in the context of this thesis, since the labels proposed here also deal with more or less the same factors. Koenigsberger, Ingersoll et al. (1974) classifies tropical climates into a further six categories: warm-humid, warm-humid island, hot-dry desert, hot-dry maritime, composite or monsoon, and tropical upland.

Lam, Tsang et al. (2005, and references 2,4, and 5 therein) mention three systems that divide China into six, ten, and five major climatic types respectively. The first of these systems, dating from 1958, used accumulated temperature or the “total sum of the temperatures in the period during which the temperature is greater than or equal to 10°C”, aggregating temperature and duration in a manner almost identical to Degree Day. The second classification system used two additional indices – “mean annual temperature of the coldest month and the annual extreme minimum”. Only the last system, which dates from 1993, is specifically for the design of buildings. It mainly uses the average temperatures in the coldest and hottest months of the year, with the number of days of daily average temperatures less than 5°C or above 25°C acting as complementary indices. Bansal, Hauser et al. (1994) and Bansal and Minke (1995) grouped India into six climatic zones based on temperature, relative humidity, precipitation, and number of clear days per month. In their system, the criteria they defined must prevail for at least six months in order for a location to be classified in a given zone. For locations that could not meet this cut-off, a catch-all category is defined – “composite”.

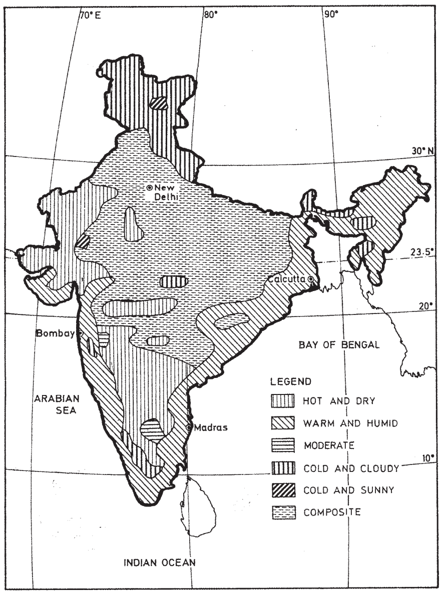


Figure 2.3 – *A map of climate zones for India, in Bansal and Minke (1995). Reproduced with permission from the first author and publishers.*

2.2 Future Climate, Climate Change, and Buildings

Current practice is to model the energy performance of buildings based on a representation of typical climate based on historical data, as discussed above. However, using historical data to predict future performance in the face of expected climate change exposes buildings to significant risks. de Wilde and Coley (2012, and reference 2 therein) mention (possibly fatal) overheating, Heating, Ventilation, and Air Conditioning (HVAC) capacity mismatch, increased wind loads, flooding, etc. In this section, we discuss existing research on *adaptation* to climate change, since any work on energy efficiency or high performance buildings may be classified as climate change *mitigation*. Morton and Bretschneider (2011) carried out a survey of a large engineering firm in the UK[1] to gauge attitudes in industry toward climate change. They state that, by and large, participants showed a high level of concern and awareness of climate change. Current practices were perceived as inadequate, and most expressed a need to develop new ways of tackling this issue. The most important finding, in our view, is that the participants focussed almost exclusively on climate change *mitigation* rather than *adaptation*. The UKCIP02 and UKCP09 climate projection reports spurred a number of impact studies in the UK. Some of these are discussed here. The comprehensive COPSE[2] project has examined the impacts of climate change on the built environment through an extensive study for the UK. Issues looked at include the creation and quality of future climate inputs and the interaction of climate change and comfort, noise, Urban Heat Islands (UHIs), and energy.

2.2.1 Future Climate Inputs

Efforts to incorporate climate change into building design and simulation have focussed on two themes: how to generate future climate files and how to adapt buildings to unknown future climates. Guan (2009) divides work on predicting future climate data into two categories: one that relies on historical data and the other on fundamental physical models. The historical data category includes extrapolation, imposed offset methods, and stochastic generation. Extrapolation is the straightforward extension of recent historical trends into the future, usually used with a simplified energy calculation like the Degree Day method, e.g., Cox, Drews et al. (2015). Imposed offset methods include the 'morphing' procedure put forward by Belcher, Hacker et al. (2005), and others that deal with two variables – temperature and humidity – by postulating some assumption on the future values of relative humidity, e.g., Guan, Yang et al. (2005). Finally, stochastic generation would include the weather generators, like Eames, Kershaw et al. (2011). The alternative to these data-based approaches is the use of numerical climate models. For example, using Global Climate Models (GCMs) to generate local weather files through computational simplifications like downscaling

[1] May refer either to the United Kingdom or the British Isles in various publications.

[2] Full name: Coincident probabilistic climate change weather data for a sustainable built environment.

or Regional Climate Models (RCMs). Guan (2009) themselves suggest a schema for generating future climate files that is a mixture of imposed offset (i.e., linear transforms of current time series) and a more detailed 'diurnal modelling method' (using current diurnal patterns with expected future distribution characteristics like daily minimum and maximum). The method proposed in this thesis could be characterised as a combination random-offset approach: we apply current diurnal patterns with noise to future low-frequency series.

To simulate the future performance of a building, and thereby have an explicit estimate of some performance parameter under estimated future conditions, a 'future weather file' is needed. That is, a time series of projected, physically viable, values. Jones, Harpham et al. (2010) list some conditions for a future time series to be useful:

- It must be internally consistent. For example temperatures are usually higher on dry days than on wet days, in the summer.
- The statistics of the series must be consistent with a "range of observed and projected statistics of the variables" from, say, RCMs of the area.
- They should adequately "represent extreme events such as prolonged rainfall, droughts and heat waves".

The most popular current approach is that of Belcher, Hacker et al. (2005), called 'morphing'. It is a simple solution that can be easily implemented in the context of building simulation, since it only requires one of three operations: addition (shifting), multiplication (linear stretching), or a combination of the two (shift and stretch). Shifting is applied to those variables for which an absolute change of mean is given in the climate change forecasts. Stretching works when the change to mean or variance is given as a fractional change. The combination is used when both the mean and variance of a variable need to be changed, e.g., if the forecast includes a change of minimum and maximum temperatures in addition to a change of mean temperatures. Belcher, Hacker et al. (ibidem) demonstrated their method for three cities in the UK. They succeeded in demonstrating the agreement of future Heating Degree Day (HDD) values calculated using their 'morphed' Test Reference Year (TRY) and Design Summer Year (DSY) files, and those from the UKCIP02 report (Hulme 2002) itself, which is the source of the climate change predictions. Jentsch, Bahaj et al. (2008) demonstrate the use of morphing to produce future 'typical year' files for the UK using morphing, while Jentsch, James et al. (2013) demonstrate the same procedure for a selection of world climates. Their work is incorporated into the *CCWeatherGen* and *CCWorldWeatherGen* tools.

Kershaw, Eames et al. (2011) report on the latest future weather generator from the UK, the UKCP09 (Jones, Harpham et al. 2010). The baseline climate for the UKCP09, like most generators and projections, is 1961-1990. Only five rainfall states are considered in the generator: "dry today/dry yesterday, wet today/wet yesterday, dry today/wet yesterday, wet today/dry yesterday and dry today/dry yesterday/dry day before" (Eames, Kershaw et al. 2012a). In addition, "the use of the three-day dry sequence allows for the prediction of heat waves" (ibidem). Upon calibration, "change factors are applied to [recorded data to] generate the future precipitation". All "... other variables are created using mathematical and statistical relationships with daily precipitation and the previous day's weather" (Eames, Kershaw et al. 2011), thereby preserving the 'Inter-Variable Relationships' (Jones, Harpham et al. 2010). The UKCP09 generator outputs 100 runs of 30 years each, from which the authors constructed 100 reference years. Eames, Kershaw et al. (2011) use the percentiles of monthly mean Dry Bulb Temperature (TDB) to create reference years tied to certain percentiles from this set of 3000 years. "This means that the median... January is combined with the median February, March, etc.", the 90th percentile January with the 90th percentile February, and so on. Eames, Kershaw et al. (2012a) compared morphed weather data with that obtained from the UKCP09 (stochastic) generator using "reference weather files". They found that for "... each location the morphing procedure systematically produces warmer minimum temperatures and cooler maximum temperatures", though they clarify that this is probably due to the fact that the UKCP09 generator produces a hundred times more files than the morphing procedure. Naturally, morphing produces the "exact same weather patterns" while the stochastic generator produces unique patterns for every year. However, the stochastic generator may also produce unreasonable values. They conclude their evaluation by saying that, while both types of data can be used for climate change studies, the morphing procedure should be used with caution for overheating studies because it does not reproduce extreme temperature well.

The reader is directed to Eames, Kershaw et al. (2011) for details of the generating procedure for the random files (the PROMETHEUS project), to Jones, Harpham et al. (2010) for details on the *UKCP09 Weather Generator*, and to Levermore, Courtney et al. (2012) for *future* TRY and Design Reference Year (DRY) files for the UK. Finally, see Mylona (2012) for a review of existing work on producing future weather files for simulation in the UK. A comprehensive study by Levermore, Courtney et al. (2012, chpt. 2) derived new future weather files for the UK, the DRYs, which are future extreme years selected based on combination of temperature, solar radiation, and humidity. Jentsch, Eames et al. (2015) also propose the creation of near-extreme Summer Reference Year (SuRY) as an alternative to the DSY by adjusting the TRY of a given site with change projections. This, they say, helps maintain a link to the TRY while incorporating climate change. An alternative to this is the DRY proposed by Levermore, Courtney et al. (2012), which is based on extracting the closest months to a particular percentile of mean monthly TDB from the 3000 years generated by UKCP09. Twenty samples of each month are initially extracted and then one is selected based on its 'typicalness' for that set of synthetic months, using the Finkelstein-Schafer statistic (FS statistic) (Finkelstein and Schafer 1971; Wilcox and Marion 2008) familiar to users of Typical Meteorological Year (TMY) files (see Section 2.6.1 for an explanation of the FS statistic and its use to create typical years).

Crawley (2007) tested the loads generated by future typical and high-low years on a test building in three locations – extreme cold, mid-latitude, and tropical – in the Americas. The study is unique in that it considers both urbanisation and climate change simultaneously. The high-low years were selected based on Degree Days, which is usually assumed to be a good proxy for climate stress on a building. Echoing Kershaw, Eames et al. (2010), they found that this was not particularly reliable. That is, "... selecting weather data based on single, simple climate descriptors such as degree days would not guarantee the lowest or highest energy for the period of record" (Crawley 2007). An expanded study (Crawley 2008) involved 25 locations in 20 climate regions. In this study, climate change predictions are applied to individual time series from typical weather files. Each current temperature value is modified with a mean change factor. In addition, the values are also adjusted to match the new predicted diurnal temperature ranges. Two values of solar radiation were calculated from current and future cloud cover values, and the current solar radiation values were multiplied by the ratio of the future values to the current ones.

Zhu, Pan et al. (2016) propose a method to generate future climate files for China based on morphing. However, the monthly-resolution climate signal is generated by fitting a periodic term to historic climate trends in China. They find that the long-term climate change has signal has a dual periodicity – century-scale (200-600 years), and decade-scale (40-80 years). This is similar to our approach (Chapter 3), except at very different time scales. The advantage of this strategy is that they are able to introduce some randomness in the forecasts. Cox, Drews et al. (2015) propose 'simple future weather files' created by applying a low-resolution climate change forecast (difference of temperature) to an existing typical year file. They found that the prediction of the heating degree days for a building in Copenhagen did not change significantly between three modes of applying climate change predictions: a single value for the whole year (low resolution), one value per month (medium), and one value per hour (high). We find it difficult to see how their example would be applicable in a climate that was not as strongly dominated by one season as Copenhagen, and with such a small expected change in temperature. The authors found that unsurprisingly, by and large, cooling energy need rises while heating falls. “In cold climates, the net change to annual energy use due to climate change will be positive – reducing energy use on the order of 10% or more. For tropical climates, buildings will see an increase in overall energy use … with some months increasing by more than 20% … Temperate, mid-latitude climates will see the largest change but it will be a swapping from heating to cooling, including a significant reduction of 25% or more in heating energy and up to 15% increase in cooling energy” (Crawley 2008). In addition, a building built to recent standards (ASHRAE 90.1-2004) performs worse than a “low-energy” buildings. A “developing case”, i.e., a building without energy codes, performs worst, though it is not discussed in full (ibidem).

2.2.2 Impact of Climate Change

Research opinion on adapting buildings to climate change is also divided into two approaches: those favouring passive design and those favouring adaptive systems. Several studies have attempted to analyse the expected impacts of climate change, e.g., references 35-42 in de Wilde and Coley (2012). There is some degree of disagreement regarding the extent of this expected impact on different building types: a mix of studies found passive buildings as being more resistant than high-tech adaptive ones, and other studies reported the opposite. There is broad agreement however, that energy use for heating will reduce in general (particularly in temperate climates), with some of it shifting to cooling. The results of climate change impact studies and efforts to predict future weather data are both bedevilled by an obvious problem – validation. Essentially, *there is no way to validate future environmental impacts without actually observing their occurrence*.

There have been a large number of case studies on the potential impact of climate change on individual case studies. Presenting the results of each individual study will be tedious, so we present the broad trends predicted. Most authors concur that cold climates will see a slight to large decrease in heating need, warm climates will see medium to large increases in cooling need, and temperate climates will see both. The worst outcome will be the creation of cooling demand and overheating risks in climates where this has not been an issue so far. The effect of urban heat islands exacerbates overheating trends (Crawley 2008). Another theme that comes through is the switching of energy sources to account for the changed needs. While the creation of cooling demand in hitherto heating-only climates increases energy use, the possible shift of some load from heating to cooling opens up the possibility of using passive measures and clean electricity.

A recent report by CIBSE (2005) highlighted several challenges to the built environment in the UK from anticipated climate change[3]. These included failure of ventilative cooling and the need for advanced passive cooling in future urban dwellings, the need for mixed-mode cooling in offices and school, and other strategies such as night flushing, reduction of internal gains, advanced solar shading, etc. They found that overheating in already-warmer regions (southern England) is likely to be several decades ahead of cooler regions (e.g., northern England and Scotland). Other impact studies based on the UKCIP02 and UKCP09 projections include Collins, Natarajan et al. (2010), de Wilde, Rafiq et al. (2008), de Wilde and Tian (2010), Jenkins, Patidar, Banfill et al. (2011), Jenkins, Patidar and Simpson (2015), Jenkins, Gul et al. (2013), Kershaw, Eames et al. (2011), Levermore, Courtney et al. (2012), Natarajan and Levermore (2007), Patidar, Jenkins, Banfill et al. (2012), Patidar, Jenkins et al. (2011, 2012), Ren, Shankland et al. (2012) and Tian and de Wilde (2011a,b). These are in addition to the case studies included in nearly every proposal for new future climate files discussed before, and the uncertainty discussions in Section 2.7. Tian and de Wilde (2011a) tackle the problem of reducing computational load by picking specific files. They compared three methods: using FS statistic statistics to pick typical files, simple linear regression between Degree Days and loads, and more complicated regression models composed of Degree Days, U-values, Solar Heat Gain Coefficient (SHGC), and Internal Heat Gains.

Patidar, Jenkins, Banfill et al. (2012) propose a modified vector Auto-Regressive (AR) model for future summer indoor temperatures based on seven weather parameters. The authors don't use this term though, preferring to call it a "simple linear model". The number of lags considered in the model is rather high at 72 hours (readers will see that our models, described in Chapter 3, consider 4 or fewer lags). The prospect of calculating 504 regression parameters (72×7) motivates the use Principal Component Analysis (PCA) to reduce the dimensions down to 33. Different regression parameter values are calculated for each Principal Component (PC) in May-June, July-August, and September-October. A second level of regression model is proposed for two 'adaptations' – window opening and external shading. The model prediction results presented are impressively accurate. However, the model does not consider and internal heat gain factors, and it is only tested on the bedroom temperature of a simple dwelling.

Studies from other regions and datasets include Switzerland (Frank 2005), USA (Craw-

[3] On the basis of a quantitative study for London using UKCIP02 forecasts.

ley 2008; Scott, Wrench et al. 1994), China/Taiwan (Huang and Hwang 2015; Wan, Li et al. 2012; Zhu, Pan et al. 2016), Singapore (Law 2013), Australia (Guan 2012), The Netherlands (Albers, Bosch et al. 2015; van Hooff, Blocken et al. 2015), Germany (Rannow, Loibl et al. 2010), Portugal (Aguiar, Oliveira et al. 2002), etc. Pyke, McMahon et al. (2012) developed two indices – a Climate Sensitivity Index and a Climate Adaptation Opportunity Index – and applied them to "potentially sensitive and adaptive practices" from LEED®-certified projects across the world. The first index highlights risks while the second indicates plausible adaptation strategies. They are both based on assigning values of strength/importance, duration of action, and reliability of control to each existing LEED® credit. All of these used building simulation or characteristics in some form. We have not included studies that drew on broad relationships between climatic variables, usually temperature, and aggregate energy or electric consumption.

There has also been some criticism of the excessively energy-centric approach to climate change adaptation (de Dear 2006; Humphreys, Nicol and Roaf 2016; Nicol, Humphreys and Roaf 2012; Roaf, Crichton et al. 2009). McGilligan, Natarajan et al. (2011) propose a new metric, the Adaptive Comfort Degree Days, that use the adaptive standard of comfort to determine future cooling demand. This is an important limitation in the case studies used in our work as well. We always use a dead-band approach to thermal comfort – assuming simply that if the temperature is below 18°C heating will be required, and if it is above 26°C, cooling will come on.

2.3 Simulation for Energy-Conscious Design

There are a number of publications that seek to offer general guidelines for energy-conscious design. Many of these include instructions on calculation techniques for specific tasks, e.g., Clarke (2001) for building physics calculations, McQuiston, Parker et al. (2005) for analysis and design of HVAC systems, Hodge (2010) for alternative energy systems, Grondzik, Kwok et al. (2011) for general building system analysis and design, and Krarti (2000) for building energy audits. Publications supported by professional organisations include *ASHRAE Advanced Energy Design Guide for Small Office Buildings*; *ASHRAE Advanced Energy Design Guide for Small to Medium Office Buildings*; *ASHRAE Standard 140-2011*; *The ASHRAE Guide for Buildings in Hot and Humid Climates*; *The MINERGIE Standard for Buildings*; *An Architect's Guide to integrating energy modeling in the design process*; *The RIBA Guide to Sustainability in Practice*. These tend to be a mixture of prescriptions, like recommended U-values for windows, and standardised methods for calculations, e.g., the ASHRAE climate classification system. Texts from individuals or groups of authors familiar to us include

Athienitis and Santamouris (2002), Bansal, Hauser et al. (1994), Donn (2009), Fathy, Shearer et al. (1986), Givoni (1976, 1989, 1992), Hausladen, de Saldanha et al. (2012), Koenigsberger, Ingersoll et al. (1974) and Olgyay and Olgyay (1992). Publications by professional bodies tend to be written in code-like language[4], while books by individual authors or groups tend to be written as narratives. Their style and content often differ based on their audience (for example, architects vs. HVAC engineers). The work contained in these publications is, by and large, meant to catalogue "shared knowledge".

2.3.1 Comfort Models for Simulation: A Brief Overview

> ... achieving thermal comfort pre-dates by thousands of years the development of the theory of heat exchange.
>
> *Standards for Thermal Comfort*,
> Humphreys, Nicol, Sykes et al. (1995, chpt. 1)

As we have discussed previously (Chapter 1), this thesis treats comfort as the basis for energy-conscious design. However, this thesis does not contribute anything to comfort models or occupants' concerns, so we will only briefly discuss the literature related to comfort measurement, modelling, and quantification. We discuss only *thermal* comfort, not related topics such as visual and acoustic comfort, or ventilation-related issues.

Comfort models, based on heat exchange between the body and its surroundings, have been developed over several decades. Readers interested in the development of thermal comfort research are directed to Carlucci and Pagliano (2012), Hensen (1990), Hoof (2010) and Nicol and Wilson (2010) for overviews. Broadly, thermal comfort theory can be separated into static and adaptive models. The former assumes that humans prefer temperature and humidity to stay within a relatively narrow band, regardless of outdoor conditions or time of year. The latter expands the comfort bands and allows them to vary based on the outdoor temperature, down and up to some limit.

The work of Fanger (1970) laid the foundations of the static thermal comfort models. Readers wishing to examine the details of this model are directed to *ASHRAE Handbook: Fundamentals*. The models are based on assuming some level of clothing (for

[4] Code as in laws, like those for structural safety, not instructions for computers.

insulation) and activity (for metabolic heat production). They are intended to deliver a "neutral" environment, where the occupants are neither too hot nor too cold, as assessed by a Predicted Mean Vote (PMV) and Percentage People Dissatisfied (PPD). The PPD cannot go below 5%, indicating that it almost impossible to satisfy every occupant of a building.

Givoni (1992) and Olgyay and Olgyay (1992) represent a pre-Fanger perspective on comfort, particularly in hot climates, and the development of the bioclimatic chart. The development of the *adaptive* comfort model (e.g., the one used in *ASHRAE Standard 55-2010*) may be traced through the work of de Dear and Brager (1998, 2002), Humphreys (1978), Nicol (2004) and Nicol and Humphreys (1973, 2002, 2010); and reviews like Halawa and van Hoof (2012). For details on principles, practices, case studies, and a philosophical justification for adaptive comfort see Humphreys, Nicol and Roaf (2016) and Nicol, Humphreys and Roaf (2012). Some authors have worked in specific contexts, e.g., Fathy, Shearer et al. (1986), Indraganti, Ooka et al. (2014), Kwong, Adam et al. (2013) and Manu, Shukla et al. (2016) (hot and/or humid climates like Egypt and India).

Newer work delves into the uncertainty of comfort calculations, and probabilistic assessments (de Wit 2001; Sulaiman and Olsina 2014). One of our proposals for future work (Section 5.3.3) includes the exploration of how comfort models interact with climate to contribute to uncertainty. Finally, see Humphreys, Nicol, Sykes et al. (1995), Law (2013) and Tuohy, Roaf et al. (2010) for future trends and standards.

2.3.2 Buildings as Systems to be Simulated

This section discusses simulation, particularly how it relates to the work presented in this thesis, after Iaccarino (2008).

First, one defines the system of interest and performance measures. This is followed by a geometrical characterization of the device, its operating conditions, and the physical processes involved. In our application, the average user is not going to examine the fundamental heat transfer processes for every simulation, but we expect that they are at least know which physical processes they are modelling. The formal next step is a "formulation of a mathematical representation of the system, ... governing equations and the phenomenological models ...", usually set in the building simulation software. The relative importance of inputs must be quantified next, with reference to the response. Iaccarino (ibidem) point out that the "system response of interest is a fundamental aspect of this phase". Simply put, if the user does not know what they are designing for, the use of simulation could create more confusion. Design does, sometimes, work without a defined goal. In such a case, it is probably not advisable to spend too much time coding simulation models.

Nominally speaking, one would input the "precise geometrical definition" of the system (building) at this point. Practically, simplification is unavoidable (and, in fact, desirable), because several geometrical features are of limited importance in a thermal analysis. For example, the precise shape and arrangement of furniture is not important unless one is concerned with glare from their surfaces, the details of the structural or foundation system are not important unless one suspects the presence of thermal bridges, and so on. In our experience, this is a remarkably difficult step, in teaching and practice because it is highly context sensitive and requires 'expert knowledge'. How does one communicate what effect each simplification had or will have without testing it? In which case, the element has been modelled and the effort put into doing so will be wasted. The user also has to introduce "artificial boundaries" to keep the scope of the analysis manageable. Once a "well defined mathematical representation" is in place, the continuous, differential system must be discretised in space and time. The spatial representation of the system is split up into discrete chunks to create an interconnected grid of nodes (finite elements). These nodes, in thermal simulation, represent temperatures. Since differential equations must be solved practically by numerical methods, becoming *difference* equations, the evolution of the system over time is also in discrete steps (finite difference).

2.3.3 Simulation: Usability Issues

Performance simulation is hardly the backbone of design, despite the ubiquitousness of computers. In fact, it would not be unreasonable to say that simulation is still regarded as something of a dark art, with dozens of fiddly knobs and fudge factors that do not make sense to the non-specialist. Lam, Huang et al. (2004) state that most dynamic simulation tools are used for "design verification and to meet building code requirements at the end of the design phase", instead of support and feedback during the design process. While those findings were published more than a decade ago, recent work, e.g., Attia, Hensen et al. (2012) and de Souza (2012, 2013), does not show a significant improvement in attitudes toward performance simulation as a design tool. The popularity of green-building or sustainability certification schemes[5], and the adoption of energy codes[6], creates a significant demand for building simulation

[5] Leadership in Energy and Environmental Design (LEED®), Building Research Establishment Environmental Assessment Methodology (BREEAM®), Green Rating for Integrated Habitat Assessment (GRIHA®), etc.

[6] *ASHRAE Standard 90.1-2010*; *ASHRAE Standard 189.1-2011*; *International Energy Conservation Code*, etc.

because of the prevalence of energy-specific criteria in the rating/evaluation systems. However, the evaluation criteria often only ask for late-stage energy analyses and do not concern themselves with the appropriateness of the design. That is, they work from a 'baseline' building usually defined on the same layout and massing, and then work their way towards more efficient envelope and system choices.

Some of the unpopularity of dynamic simulation tools is due to inappropriate design of the user interface or even the core software. Alsaadani and de Souza (2012), de Souza (2012) and de Souza and Knight (2007) make some excellent points on the difference in thinking between tool designers and users, which causes the tool design to diverge from the requirements of the users. Papers on usability and acceptance have presented a gamut of reasons both for and against the adoption of energy simulations in design, based on surveys of practitioners and academics. We present the following non-exhaustive list of pros and cons based on Attia, Hensen et al. (2012), de Souza (2012) and Wong, Lam et al. (2000). Reasons in favour include enhanced understanding of the impact of design choices on final performance, increased speed of iteration, and better confidence in the design. Arguments against using simulation tools included extra cost and effort with little resultant recognition from clients; tight project schedules and budgets; lack of in-house skills, or local training and support from vendors; steep learning curves; the very extensive data input necessary (especially because of incompatibility with CAD software); and, lack of knowledge about the physical principles. Finally, interpreting the results, especially in the light of uncertainty, is mathematically challenging.

This problem is compounded by the fact that building performance simulation specialists cannot agree on standards of inputs, outputs, and performance. The BESTEST procedure, codified in *ASHRAE Standard 140-2011* (Judkoff and Neymark 1995, 2006), is an important step in this direction. Attia, Hensen et al. (2012) identify a host of issues with comparing and standardising Building Performance Simulation (BPS) tools. They find that there are no uniform definitions of "tool requirements and specifications", no clear methods to compare different tools, and "... no common language to describe what the tools could do" (Crawley, Hand et al. 2008). However, the authors find a substantial amount of literature attempting to rank and compare BPS tools. For example, Crawley, Hand et al. (2005, 2008) listed a series of studies comparing building "energy programs", and conducted a comparative survey themselves – though it was based on vendor-supplied information with "... limited peer review". Perhaps the most interesting recommendation in their documents is that of a 'living' document that catalogues the capabilities and performance of various software regularly. From the literature, it seems that there is a lack of consensus among practitioners about the utility of these tools and among specialists about the standards. In a paradigm of prescriptive design codes and “a fragmented building delivery process that does not routinely include quantifiable assessments of design options”, it is understandable that rigorous simulation falls by the wayside Wong, Lam et al. (2000).

There are so many tools available for energy modelling of buildings now that it is quite reasonable to assume no professional can hope to be familiar with more than a small subset, let alone have any degree of proficiency in them (a list online, *BEST Directory: Building Energy Software Tools*, includes more than a hundred tools). Donn, Selkowitz et al. (2012) suggest, quoting findings from a usability study conducted for the tool COMFEN (Hitchcock, Lee et al. 2008), that any design guidance tool must try to answer certain question about performance and the likelihood of getting this performance. We summarise the questions here [with some of our own additions in square brackets]. The themes raised by these questions show up throughout this thesis.

- What are the expected [energy] costs of operation?
- What variations of comfort [or energy use] are expected? Both with the passage of time and in the different parts of a building.
- What are the risks for comfort and cost [and what is their likelihood of occurrence]?
- What sort of interaction is expected between the building systems and users?
- What is the likely impact of future climate change [on building performance]?

PRÉCIS

- Building simulation is widely considered to be cumbersome or complicated.
- The use of simulation is not adequately rewarded or recognised.
- There are few standards to judge simulation software, and little agreement on how to evaluate them.
- The precise results from numerical simulation are not well suited to a decision-making process that is often qualitative.

■ For a variety of reasons, the use of simulation in *design* is neither widespread nor deeply integrated.

2.3.4 Using Simulation for Early- and Late-Stage Design

The 'early stage' of design is a commonly used expression for a somewhat vague concept. While it is not clear where this stage ends, it is generally accepted to refer to that part of the design process when initial ideas are explored and project requirements finalised in an adaptive-iterative process (Lam, Huang et al. 2004; Mahdavi and Lam 1993).

This thesis does not concern itself with the very earliest phases of design, when the client's brief is being translated into a design framework, site arrangement, and initial concepts. The 'system', i.e., the building, is too vague, or ill-defined, at this point to be usefully probed with simulation. In any case, the definition of what the early and late stages are, is itself nebulous. There seems to be a consensus that the early design phase is when decisions have most impact on energy use, which is intuitive. From the perspective of dynamic thermal simulation, the distinction between early- and late-stage design is an epistemic one. It is difficult to simulate a system with vague descriptions, unless one is willing to accept that most of the inputs have huge uncertainties, some of which may be quantified with prior knowledge. There are limits to knowledge at each stage of the design: the level of detail required to accurately model a building's energy usage is simply not available from the start. For example, when one is working on the initial massing of a building, one is not concerned with the U-value and placement of the windows – though without any knowledge of the impact these later decisions will have on the final energy design of the building. One may have ideas about what values are *likely*, as a sort of Bayesian *prior*, but the value is not fixed until a design decision is made, or even until construction. This means that using simulation at all stages may not be practical but *could be useful*.

There is some recent work on assessing the impact of early design decisions using simulation through simplification and the use of defaults, e.g., Asadi, Amiri et al. (2014), Attia, Gratia et al. (2012), Augenbroe (1992), Carlos and Nepomuceno (2012), Gervásio, Santos et al. (2014), Lam, Huang et al. (2004) and Pranovich, van Wijk et al. (2003). The work of Granadeiro, Correia et al. (2013), Hygh, DeCarolis et al. (2012), Jin and Overend (2014), Ochoa and Capeluto (2009) and Thalfeldt, Pikas et al. (2013) seeks to address the design of envelopes in particular. Santos, Martins et al. (2014) and Schlueter and Thesseling (2009) look at the prediction of operational energy use at the early design phase. Of these, we will discuss the approaches of Asadi, Amiri et al. (2014) and Hygh, DeCarolis et al. (2012) in more detail later, since they use regression/numerical methods. We have also published some work with collaborators on the sensitivity of predicted energy use to an early design parameter – the urban geometrical factors

(Nault, Rastogi et al. 2015). We direct the reader to a colleague's recent work (Nault 2016), where they review the development of tools for the early-design phase and the concepts underlying them.

2.3.5 Optimising for Energy and Comfort

There is nothing like looking, if you want to find something.
You certainly usually find something, if you look,
but it is not always quite the something you were after.

J. R. R. Tolkien (Thorin II 'Oakenshield'),
in *The Hobbit*

Optimisation is a promising, if elusive, tool for design. Optimisation, defined formally, works only in a well-defined context of quantified inputs, outputs, and objectives. It is telling that, in a survey of practitioners about the use of optimisation, Attia, Hamdy et al. (2013) found that "all interviewees (28) chose energy as the most used optimization objective, while 64% (18) chose cost" – probably two of the most easily quantifiable aspects of design. In addition, "70% of all interviewees do multi-objective optimization versus 30% who do single objective optimisations". Evins (2013) found that about half the publications used single-objective optimisation, while 40% used "full Pareto multi-objective optimisation".

The number of publications that seek to optimise a certain sub-system or particular performance aspect is huge[7]. A cursory search through our archives gives several examples: Gagne and Andersen (2010), Madeddu (2011) and Vartiainen, Peippo et al. (2000), for daylight and associated systems; Ascione, Bianco et al. (2015), Caruso and Kämpf (2015), Ellis, Griffith et al. (2006), Glassman and Reinhart (2013), Murray, Walsh et al. (2014), Pont, Shayeganfar et al. (2013), Ramallo-González and Coley (2014) and Yang, Li et al. (2014), for energy-based goals; Hoes, Trcka et al. (2011) and Ramallo-González, Blight et al. (2015) for robustness considering user behaviour; and, Coffey (2012), Lazos, Sproul et al. (2015), Lindelöf (2007) and Mahdavi and Mahattanatawe (2003), for building controls. The work of Andersen, Kleindienst et al. (2008), Gagne (2011) and Gagne, Andersen and Norford (2011) proposes an expert system, a sort of hybrid approach involving a knowledge database, human judgement, and optimisation. This is an optimisation-assisted approach in which the user input changes the trajectory of the process at every interaction. In this

[7] To the best of our knowledge, the review of 165 publications by Attia, Hamdy et al. (2013) has not been trumped.

case, the 'human' is not taken out of the 'loop'. Several of these publications address more than one aspect, of course, so this classification is somewhat loose. We did not review the literature on optimisation of HVAC systems, which is extensive but not intimately connected to building design. The interaction between systems and building construction/occupancy is non-linear and complex, but in the case of this thesis we are chiefly concerned with the building and its intrinsic responses. For more thorough surveys of the use of optimisation techniques in sustainable building design, we direct the user to the exhaustive reviews of Attia, Hamdy et al. (2013) and Evins (2013). The former is a general review of the use of optimisation for *any* building-related design problems, while the latter works in the context of integrating optimisation tools for the design of Net Zero-Energy Buildings (NZEBs).

A common thread through much of the literature is that the so-called 'direct search' algorithms do not perform satisfactorily. "Direct search covers methods that compare trial solutions with the best found so far, ... using results so far [to determine] the next trial" (Evins 2013). Consequently, the development of optimisation in design has tended toward heuristic (also called meta-heuristic) methods like genetic algorithms and evolutionary algorithms, since these methods work well for design problems that are "discontinuous, non-differentiable, stochastic, or highly non-linear"[8] (The MathWorks, Inc. 2015). Heuristic methods differ in one crucial aspect from classical methods: "they do not guarantee to arrive at the true optimum, but offer an efficient method that has a high probability of finding the optimum or of getting close to it" (Evins 2013). Attia, Hamdy et al. (2013) find that this is not necessarily a bad thing for design. Their survey of users (practitioners, etc.) finds that many regard the concept of a 'true optimum' nonsensical, given that there are large uncertainties in the process and that the objectives themselves may change over time. Optimisation is also seen as an exploration rather than a strict search. Genetic/evolutionary algorithms mimic natural selection and randomly permute a set of 'genes' or components to create populations of solutions at each step, say a population with varying combinations of Thermal Mass and U-value. They select the best performers at each step, allow the individuals to mutate and cross-over for the creation of a new generation (e.g., new combinations of Thermal Mass and U-value not tested before). The 'unfit' solutions are discarded from each generation, and only the best performers become 'parents' for the next step. Attia, Hamdy et al. (ibidem) report that the use of evolutionary algorithms is seen to be particularly promising in highly constrained problems related to envelopes and systems.

[8] These are the qualities of the objective or cost function that is being evaluated. Incidentally, this is more or less why classic linear models do not perform well in our own work on creating emulators (Chapter 4).

Two classic problems that occur in the application of optimisation to building systems are "uncertainties in simulation model input, and long computational times" (Ramallo-González and Coley 2014). The issue of uncertain inputs in optimisation has been been addressed in recent work on 'robust optimisation', e.g., Aïssani, Chateauneuf et al. (2015), Marijt (2009), Ramallo-González, Blight et al. (2015) and Van Gelder, Janssen et al. (2013). Robust optimisation involves the minimisation of an objective function under a range of input conditions, as opposed to a deterministic values (like typical weather). Some recent work, including by us, addresses robustness specifically to a changing climate, e.g., Chinazzo (2014), Chinazzo, Rastogi et al. (2015a,b) and Nik, Mata et al. (2015), though these publications do not use formal optimisation routines. There are two ways to address the issue of computational time: simplified models or emulators. These are discussed in more detail below (Section 2.5), under *data-based* and *physics-based* methods (emulators vs approximate models).

Recent work proposes the use of "adaptive" algorithms. That is, algorithms that do not waste time in sampling the whole design space to the same granularity/depth. Much efficiency can be gained by the use of adaptive search if: not all variables are of interest, and the user may employ domain knowledge to reduce dimensionality; not all variables reveal interesting features upon increasing resolution, i.e., the response is smooth against those variables; and not all variables have the same impact on the objective, i.e., the output shows differing sensitivity to inputs. This is a logically attractive approach, and consonant with the interpretation of the meta-modelling approach proposed in this paper for uncertainty/sensitivity analyses. For example, the use of adaptive sampling would be efficient to train emulators from expensive simulations. In this thesis, we already employ correlation and PCA to reduce the number of predictors/inputs. The work of Eisenhower, O'Neill et al. (2012) and Ramallo-González and Coley (2014) on emulator-assisted optimisation relates closely to a possible extension of this work, discussed in Chapter 5.

2.4 Uncertainty and Sensitivity in Simulation

They say that a little knowledge is a dangerous thing,
but it is not one half so bad as a lot of ignorance.

Terry Pratchett, *Equal Rites*

We are aware that this section is about to summarise/mangle several treatises' worth of philosophical arguments in a few sentences. Fundamental questions of epistemology, lit. *study of knowledge*[9], have been asked through the ages and generate much intelligent debate and journal papers. For more discussion on the nature of knowledge and knowing, we direct the reader to the nearest professional philosopher. This thesis is about the nature of inputs and outputs to building simulation, so we are more concerned with uncertainty and sensitivity affecting the validity and veracity of computer models, rather than the nature of knowing.

Uncertainty analysis is a term encompassing the set of methods that could be used in to quantify and systematise a user's lack of knowledge or assessment of the apparent randomness of a physical quantity. Uncertainty is a little more esoteric than sensitivity, mostly because there are several different interpretations on offer, none of which offer much by way of certainty. Davison (2003) discusses the various interpretations of Uncertainty Analysis (UA) and confidence intervals. In this thesis, we use a *frequentist interpretation*, i.e., one based on repeated sampling of a computer simulation using pseudo-random inputs to construct variability intervals. Sensitivity analysis is the substantially different concept of examining the relationships between inputs and outputs. However, sensitivity may also be expressed in terms of variability intervals, as in the possible values of an output given a certain range of inputs. This is especially useful when dealing with inputs whose relationship with outputs is not analytically simple. Macdonald, Clarke et al. (1999) point out that while "the most widely used methods for assessing uncertainty are borrowed from Sensitivity Analysis (SA)", a "distinction must be drawn between a sensitive parameter and an *important parameter*". If a sensitive parameter, i.e., a parameter to which the system is sensitive, is known to high degree of certainty, it "will not lead to significant uncertainty in the predictions" (D. M. Hamby, 1994, cited in ibidem). An important parameter, on the other hand, is generally so labelled because of its significant relative contribution to the output uncertainty. A disadvantage of numerical models is that they do not give explicit *functional relationships* between and among the output variables, input variables, and sundry parameters. This is the case with most building performance simulation, since analytically solvable relationships are hard to find among highly complex equations governing the effect of certain variables (e.g., material properties, environmental effects, etc.) on others (e.g., indoor temperature). This necessitates the use of uncertainty and/or sensitivity analyses to understand the response of a system to changes in the variable of interest.

Introducing UA into building simulation programs is, by and large, an ad hoc affair. The modeller assumes some distribution of the populations of different inputs and tries to sample the distributions without bias. Some literature on UA separates it

[9] "Defined narrowly, epistemology is the study of knowledge and justified belief" (Steup 2014).

into two types based on the source of the "doubt": epistemic and aleatory. The first arises from a lack of knowledge about a physical phenomenon or object. The second arises from the randomness inherent in the inputs or the system. While one could object that the 'randomness' is an illusion created by epistemic uncertainty about the inputs, system, or associated phenomena, for our purposes some phenomena are taken as random. Weather, occupant interaction, and construction errors are obvious examples. Hopfe and Hensen (2011) and Hopfe (2009) divide uncertainty into three categories – physical, design, and scenario. de Wit (2001) use a different classification – specification, modelling, numerical, and scenario. Specification uncertainty is said to arise from the incomplete or simplified representation of a building; modelling uncertainty from assumptions and simplified representation of physical processes; numerical uncertainty from discretisation and simulation in finite-precision computer systems; and scenario uncertainty from lack of knowledge about boundary or forcing conditions like weather and occupancy. Macdonald, Clarke et al. (1999) use yet another nomenclature, based on sources: abstraction, databases, modelled phenomena, and solution methods. The first two categories are both specification uncertainties, 'modelled phenomena' includes modelling and numerical uncertainties, and 'solution methods' overlaps with the numerical uncertainties group.

Macdonald, Clarke et al. (1999), Macdonald (2002) and Macdonald and Strachan (2001) mention some uses to which UA can be put in building simulation:

- "[Building] Model realism: How well (and to what resolution) does the model represent reality?" [epistemic, model uncertainty]
- "Input parameters: What values should be used in the absence of measured data?" [epistemic, specification uncertainty]
- "Stochastic processes: To what extent do the assumptions made regarding future weather, occupancy and operational factors affect the predictions?" [aleatory, scenario uncertainty]
- "Simulation program capabilities: What uncertainties are associated with the particular choice of algorithms for the various heat and mass transfer processes?" [epistemic, numerical uncertainty]
- "Design variations: What will be the effect of changing one aspect of the design?" [This we would classify as a Sensitivity Analysis (SA) rather than an UA, unless there is some doubt about whether a design option will be chosen or not.]

2.4.1 Errors vs Uncertainty

The difference between errors and uncertainty is that the former are "recognisable deficiencies", whereas the latter are "potential deficiencies due to lack of knowledge". Another useful way to think about this is that errors are a result of "the translation of a mathematical formulation" into an algorithm/code, keeping in mind that the mathematical formulation is itself an approximation of a natural phenomenon or system. These errors could occur due to numerical approximation/representation, or round-off errors, convergence issues in iterative algorithms, and even 'clerical error', mistakes in implementation (bugs). Uncertainty, on the other hand, arises from the choice of physical models and the "specification of their input parameters" (Iaccarino 2008). We now discuss the two broad categories of uncertainty in simulation (epistemic and aleatory).

2.4.2 Types of Uncertainty

2.4.2.1 Aleatory

Aleatory uncertainty is said to arise due to the inherently random or variable nature of a quantity, or the (usually unknown) system underlying it. It is "irreducible", because better information or experiments cannot eliminate it. Instead, one achieves a better understanding of the distribution of the quantity through acquisition of relevant experimental data (ibidem). The uncertainty we assign to climate falls under this category, as does user behaviour. Does that mean that we are claiming that the earth's climate system is inherently random? Not at all. However, there are no perfectly accurate models of the climate, so, from the perspective of simulation, climate input is largely random. Chapter 3 gives more details about how this randomness is modelled. The same can be said of user behaviour, which will be the subject of future work (Section 5.3.8).

It is useful to think of aleatory uncertainty in terms of *possibility*, i.e., what *could be*. The designer has no way of knowing, for sure, whether there is going to be a heat wave in the next decade. One can say, with finite confidence, that the temperature is likely to be this or that, based entirely on computer models. So, we suggest that the designer assume that the heat wave *both will and will not occur*, and plan accordingly.

Since there is no way of knowing the weather at a certain time in the future without actually experiencing it[10], robust design boils down to accepting this interpretation. A

[10] Aleatory uncertainty is irreducible.

probability distribution on temperature does not imply that, for example, a 100-year heat wave (say 37°C) returns exactly every one hundred years like clockwork. It just means that it could happen every year, but there is only a one-in-hundred chance of that. If the temperature touches 37°C every year for a decade, then clearly the definition of a 100-year event needs to be changed. Recall that this is a frequentist interpretation of probability, i.e., one based on acquired data, so it can only be formed after the event has occurred (*post-hoc*). With climate models, we can assign Bayesian priors to the probabilities of future events. In the case of the synthetic files in this thesis, we use 'knowledge' about the climate system – the time series models and climate change forecasts (IPCC 2014b) – but we do not formally update the randomly generated series based on 'new information', i.e., recently recorded data.

2.4.2.2 Epistemic

Epistemic uncertainty is the uncertainty "of or relating to knowledge": how, how much, and why do we know the things we know? It is literally the state of *not knowing*, arising either because there is lack of information or the information is un-knowable. It is easy to imagine a situation where the user does not know some parameters of a system, e.g., material properties, energy policies, etc. Epistemic uncertainty may be reduced in most cases through experimental investigation and calibration. Probabilistic approaches are of limited use because there is no justification of a chosen distribution if there is a nominal lack of knowledge. Bayesian interpretations may be more useful in this case, because the user may use expert knowledge or other inputs to justify a prior belief, like the skewed distributions of material properties proposed by Jain, Ramallo-González et al. (2014). An interesting example of reducing epistemic uncertainty using Bayesian methods in building simulation is in Lindelöf (2007). They propose a lighting controller that "builds an internal representation of the room... without user input". It also learns user preferences by analysing the overriding behaviour. Here, the controller is acquiring knowledge of two unknown parameters: the physical make-up of the room and the user's preferences.

While epistemic uncertainty is reducible with better information or calibration, it may not always be practical or cost-effective to do so. Say the simulation user is modelling a stock of city buildings, and does not know whether the designed wall assemblies were degraded, installed incorrectly, or remained the same during construction. This fact is theoretically verifiable with a good enough investigation but it is usually not practical to test the installed U-values of *every* building in a city. What is the user to assume in this case?

One option is to assume no change, and work with the designed values. This is common practice, and we will call this the *status quo* approach. Another is to survey a small sample of buildings and extrapolate the results to the whole building population by fitting a Probability Density Function (PDF). This we call the frequentist approach, and its validity depends on the representativeness or quality of the sample[11]. A third approach could be the creation of a Bayesian prior. A prior may be constructed purely from expert knowledge, or from a limited experiment, or from some database like the one proposed by Zhao, Plagge et al. (2015). One could conduct trials to determine the conditions for degradation, like faulty installation practices or post-installation wiring. A (probabilistic) physical model, that determines the probability of seeing degradation, or the magnitude of degradation, based on the installation conditions, is a prior. The model need not say that degradation *will* occur when certain installation conditions exist, just that it can occur with some probability. The two approaches which assign distributions to the U-values of walls – based on sample surveys or experiments – allow the user to quantify the impact of their lack of knowledge on the outcome. In the case of Bayesian priors, qualitative issues could be brought to bear. For example, maybe the user is looking to do a *worst-case* analysis to determine potential failures (fatal overheating).

2.4.3 Uncertainty Propagation

Not all those who wander are lost ...

J. R. R. Tolkien,
The Lord of the Rings: Fellowship of the Ring

A probabilistic approach to uncertainty would imply the defining of PDFs for inputs of interest. This uncertainty in inputs is *propagated* through a system to obtain uncertainty of outputs. There are several approaches available, "from sampling based approaches, e.g., Monte Carlo (MC), to more sophisticated stochastic spectral Galerkin approaches" (Iaccarino 2008). We will discuss two different kinds of approaches – internal and external – after Macdonald (2002). Internal approaches solve the model equations stochastically, while external approaches sample from the distribution of inputs to re-simulate the model several times. For an overview of both, we refer the reader to Helton, Johnson et al. (2006), Iaccarino (2008) and Macdonald (2002).

[11] Recall that using certain formulae for mean and variance assumes that the population follows the distribution on which those formulae are based.

In this thesis, we use a ‘sampling’ technique: Monte Carlo simulation. This method is “simple, universally applicable, and does not require any modification to the available (deterministic) computational tools”(Iaccarino 2008). With MC, low-order statistics, like expected value (mean) and variance, may be well-estimated with a small number of samples, but “... a prohibitively large number of realizations may be required to accurately estimate responses that have a small probability of occurrence” (ibidem). One can see that truly random sampling could easily lead to a situation where there are a lot of inconvenient “clusters and holes in the samples” (ibidem). Systematic sampling approaches exist to address this issue, e.g., stratified sampling and Latin Hypercube Sampling (LHS). Stratified sampling is when samples are taken uniformly across the entire input range, which has been divided into equally-spaced disjoint sub-ranges. In LHS, the sampling range is divided into M intervals of “equal marginal probabilities”, and one sample each is drawn from the sub-ranges (McKay, Beckman et al. 1979).

Hopfe (2009) link Uncertainty Analysis (UA) to assessment of risk, which in the context of building simulation amounts to acknowledging explicitly that a given design may cause some performance metric to exceed its tolerated limits. They treat Sensitivity Analysis (SA) as the process of identifying the “most sensitive parameters”. These are common threads running through most of the literature on uncertainty and sensitivity quantification in buildings. We feel that framing UA in terms of “risk” is a sound approach to taking it out of the research framework into practice. Engineers and designers will be on board with the idea of mitigating risk, or otherwise planning for it, because of the potential to directly affect the comfort and well-being of users. Ditto for sensitivity of outputs to different design parameters. It can potentially enhance the focussed targeting of retrofit measures and other interventions. In the work summarised below, the reader will notice that there has been far more work using ‘external’ methods than ‘internal’ ones, although some work on quantifying model uncertainty is hard to classify in either of these categories. For one, the former is easier. Secondly, not all inputs can be easily rendered into the web of differential equations that is a building simulation program. The reader is directed to Macdonald (2002, chpt. 3) for an overview of Uncertainty Quantification (UQ) techniques, both internal and external, and Macdonald (ibidem, chpt. 4) for an overview about uncertainty in building simulation. See Hopfe (2009, chpt. 3) for a discussion of uncertainty and SA for design support. The user is also invited to examine the Georgia Tech Uncertainty and Risk Analysis Workbench (GURA-W), whose goal is to promote flexible and automated exploration of uncertainties in building simulation inputs (Lee, Sun, Augenbroe et al. 2013). It contains a database of “uncertainty distributions for a variety of parameters and models” (ibidem), and can automatically identify and modify the relevant parameters in a simulation.

2.4.3.1 Internal Methods: Stochastic Models

The work of Macdonald (2002) and Macdonald and Strachan (2001) takes the 'internal' approach. However, they caution the user to the effect that the performance of the stochastic simulations is not a given, mainly due to convergence issues. "For small uncertainties the predictions made by the [internal] method agreed with those of a differential and Monte Carlo analysis. However, for larger uncertainties the method failed to produce useful bounds on the predictions" (Macdonald 2002). They suggest continuing with a mix of internal and external methods until these "limitations are overcome". This work has been implemented in the ESP-r program, though we did not test it in this thesis.

A possible line of future work from this thesis is the assessment of the suitability of internal methods for analysing the uncertainty due to weather inputs, building on the proposals of Haghighat, Chandrashekar et al. (1987) and Haghighat, Unny et al. (1985). Interestingly, their work was looking at the calculation of a rigorous safety factor[12] for design based on treating weather inputs as random, and propagating this randomness through the simulation, rather than a characterisation of the long-term behaviour of a system. The approach involves partitioning the inputs into random and deterministic components, the latter of which may or may not be time-dependent. Following this, the state equation of every node[13] is also split into a deterministic (mean) component, and a stochastic term. The stochastic term is a product of a (white) "noise term" and a "function denoting the sensitivity of the system to the noise term" (Haghighat, Chandrashekar et al. 1987). White noise is selected both for its ability to "adequately describe the random disturbance" and its analytical tractability. The latter is a crucial point for internal approaches, since the existence and uniqueness of solutions is guaranteed for white noise processes. An external procedure, like the one in this thesis, need not work only with specific distributions. In recent work, Brohus, Frier et al. (2012) proposed using stochastic differential equations to model uncertainty. For their test cases (a mechanically ventilated room and a naturally ventilated atrium), they found that the "computation time may be rather high". They also found that the impact of using stochastic methods was "modest for the dynamic thermal behaviour of buildings", but significant for air flow and energy consumption.

[12] When designing all sorts of engineering systems, the designer will typically compensate for uncertainties by over-designing the components. For example, a building with an approximately 5 kWh peak cooling demand might be serviced by the next largest available unit, say 7.5 kWh.

[13] In a finite difference network of nodes representing the building and its boundary conditions.

Silva and Ghisi (2014b) and Sun, Su et al. (2015) examined the effect of so-called "model" or "model form" uncertainties. The former looks at geometrical and material simplifications while the latter looks at the anisotropic Perez sky model as an example (Perez, Ineichen et al. 1990; Perez, Seals et al. 1987).

2.4.3.2 External Methods: Random Inputs

The literature on using external methods for uncertainty and sensitivity analyses in building simulation is extensive. We present here a very brief overview of some of the work, addressing different aspects of simulation. Weather inputs are dealt with in Section 2.7.

Booth and Choudhary (2013), Jain, Ramallo-González et al. (2014) and Sandberg, Sartori et al. (2014) assess the uncertainty surrounding the implementation of 'green' retrofit measures in the building stocks of the UK, India, and Norway, respectively. In the UK study, the main risk assessed is financial, arising from an overestimation of efficiency gains. In the India study, the risk assessed is that of the reliability of predicted loads. In the Norway study, the authors looked at the effect of uncertainty about renovation of the housing stock on long-term emissions/consumption targets. Booth, Choudhary and Spiegelhalter (2012) propose a framework for modelling uncertainties in housing stock models[14], including Bayesian calibration based on measured consumption data. Silva and Ghisi (2014a) studied the effect of uncertainty due to inputs on the energy consumption of a low-income house in Brazil. The inputs studied included several building parameters, e.g., solar absorptances; a few climate parameters, e.g., ground temperatures and albedo; and occupancy, through schedules and number of occupants. They found a "relative deviation" of 19.5-43.5% on the energy consumption for heating and cooling. Sun, Gu et al. (2014) present an uncertainty and SA framework for HVAC system sizing, focusing on rigorous safety factors like Haghighat, Chandrashekar et al. (1987). Burhenne, Tsvetkova et al. (2013) carry out a combined analysis of the risks due to simulation inputs and economic factors in predicting building energy consumption. Breesch and Janssens (2005) looked at the uncertainty in predicting the performance of night ventilation using LHS, and the sensitivity of ventilation performance to a range of building-related inputs using Standardised Regression Coefficients (SRC). Mazo, El Badry et al. (2015) examined the influence of uncertainty related to the measurements of the thermo-physical properties of Phase-Change Material.

[14] i.e., models representing types or groups of buildings in a region.

Hopfe, Augenbroe et al. (2013), Hopfe and Hensen (2011) and Hopfe (2009) present a series of case studies demonstrating the use of uncertainty and sensitivity analyses for design decision support. Their latest work proposes a framework for systematic multi-criteria decision-making under uncertainty. They argue that including the uncertainty of inputs can produce a more “well-informed analysis” but not necessarily the most straightforward one. de Wit (2001) presents a framework for quantifying the epistemic uncertainty in a range of factors considered important in determining indoor thermal comfort. They consider factors for which a statistical prior is not viable, using instead a prior obtained from expert judgement. They propose the integration of these priors into design decision-making using Bayesian decision theory. To select important parameters of study, they conducted a Sensitivity Analysis (SA) using a large set, and eventually picked only the two most important sets: wind pressure coefficients, and temperature distribution in the indoor air. See de Wit (2003) for an overview and worked example of uncertainty and sensitivity analysis with external methods.

Kim and Augenbroe (2013) present the development of a framework for the management of uncertainty in demand-side controls. They also discuss the theoretical framework for UQ in this context, leading from complete ignorance to complete certainty. Struck, Hensen et al. (2009) outfit a “conceptual design tool” with the capability to carry out UAs and SAs, and compare its performance to a “detailed design tool”. They hypothesised that if the conceptual design tool is propagating the uncertainties faithfully, then the resulting uncertainty in outputs must be higher than the detailed design tool, since the conceptual tool works at a “higher level of abstraction” (ibidem). They conclude that if UA and SA are deemed useful to increase the uptake of tool usage in the early design phase, then “detailed tools with simplified interfaces pose the most promising way forward” (ibidem). Their work makes an important point in that it is difficult to understand the influence of model and numerical uncertainties on epistemic and aleatory uncertainties. This is the same problem raised by Macdonald (2002) in their discussion on the appropriate mathematical techniques to propagate uncertainties through a system of difference equations. The problem is especially acute with commercial software, whose inner workings are not in the public domain. Parys, Breesch et al. (2012) examined the feasibility/robustness of two passive cooling systems using MC analyses for an office building in Belgium. There is a small element of climatic uncertainty as well, with the inclusion of an “extreme weather data set” (“probability of occurrence of 1-in-10 years in terms of high temperatures and high amounts of solar radiation”).

2.4.4 Sensitivity Analyses in Building Simulation

Lomas and Eppel (1992) discuss three ways of analysing the sensitivity of a building simulation program: Differential Sensitivity Analysis (DSA), Monte Carlo Analysis (MCA), and Stochastic Sensitivity Analysis (SSA). In a DSA individual parameters are changed *one by one*. Along with collaborators, we previously undertook a study like this in Chinazzo (2014) and Chinazzo, Rastogi et al. (2015a). DSA has the advantage of revealing the sensitivity of a system's response to individual inputs (Lomas and Eppel 1992). This is an advantage when only one kind of change is important. For example, adding insulation to change the U-value (or R-value) of the envelope. Fürbringer and Roulet (1995) also compare the performance of sensitivity analyses carried out using the Monte-Carlo method and factorial experimental design separately and in combination. For their case study, they arrived at the conclusion that a combination of the two is best: factorial design to get a broad overview of the domain, and Monte-Carlo at each point selected by the factorial design for an in-depth analysis at that point. Lomas and Eppel (1992) recommend a DSA for obtaining sensitivities to individual input parameter uncertainties and MCA for total sensitivities. Their purpose though was slightly different from the one envisioned in this thesis – "to assess the reliability and resolution of simulation programs for the design of passive solar buildings in the UK". The SSA proposed in their work is slightly more complicated, but conceptually appealing. It envisages the real time (i.e., during simulation) application of impulse changes in inputs to calculate changes in outputs. The sensitivities are then calculated by assuming that the input perturbations are independent white noise terms.

Tian (2013) divide sensitivity analysis methods into two categories: local and global. Global is further divided into methods based on Regression, Screening, Variance, and Meta-modelling. The authors go on to discuss each method type and its applications in building simulation. To clarify, in this thesis, when we use the term 'meta-modelling', we mean a regression-based emulator. On the other hand, Tian (ibidem) makes a difference between approaches that fit a regression equation to data from simulations, to examine the coefficients for example, and approaches that use meta-models in place of the original simulations to improve run-time. The crucial difference then, between the meta-modelling approach and all others, is that the user is really probing the *sensitivity of the meta-model, not the original system.* This is well-understood, which is why meta-modelling approaches usually involve some calibration and verification. The method proposed in this thesis is a global meta-modelling-based method, in this classification. The authors end with a discussion of some issues on the extension and improvement of sensitivity analysis, with experiments, confidence interval, better software, and acknowledgement of correlation among inputs.

Loonen and Hensen (2013) introduce the concept of 'dynamic sensitivity analysis', arguing that the aggregated metrics used in most studies, like annual energy need or peak demand, do not capture the dynamic sensitivity of a building to its constantly varying environment and operational demands. The dynamic analysis proposed in their work is relatively straightforward, though computationally intensive. Unlike in conventional analyses, where only the final result of N simulations[15] is examined, in this approach the result at every time step is stored. This way, the 'sensitivity profile' at every time step is available. The interpretation of these profiles is more complex than that of a single aggregated profile, so the authors suggest using some post-processing tasks like smoothing.

Blight and Coley (2013) looked at the sensitivity of predicted energy consumption in Passivhaus® dwellings to natural variation in occupant behaviour and household composition. Similarly Firth, Lomas et al. (2010) examined the sensitivity of predicted household energy consumption to uncertainty in the inputs to the Community Domestic Energy Model (CDEM), a stock model developed in the UK to explore household efficiency measures. They found that the influence of under-performing household energy efficiency measures on the national Green House Gas (GHG) emissions targets could be very large, and good quality control in construction and refurbishment is needed to ensure targets are met reliably. For the use of sensitivity analysis as a design tool, or as part of the design phase with specific applications, see Garcia Sanchez, Lacarrière et al. (2014), Hui (1996), Hygh, DeCarolis et al. (2012), Loonen and Hensen (2013), Nault, Rastogi et al. (2015), Sandberg, Sartori et al. (2014), Shen and Tzempelikos (2012), Smith, Aguilar et al. (2012), Struck, Hensen et al. (2009) and Tian and de Wilde (2011b). Athienitis (1989) present a 'sensitivity analysis tool' based on a discrete frequency-domain model of a thermal network model. In their formulation, the properties of components can be changed without the need to invert a matrix.

For a comprehensive overview of SA methods, the user is directed to Hamby (1994) and Saltelli (2008). For a review of sensitivity analysis in building simulation, see Tian (2013). See Kleijnen (2001) for a tutorial on experimental design of simulations for what-if or sensitivity analyses.

2.4.5 Issues: Speed and Complexity

By and large, the methods for quantifying uncertainty and sensitivity discussed above suffer from issues of complexity and computational load. Internal methods have not

[15] Using N unique input combinations.

been implemented efficiently, as discussed in Macdonald (2002). External methods, by definition, require far more simulations than a single-input run. Efficient experimental designs can go some way to alleviating that, but have not, so far, reduced computational load to the point where UQ and Sensitivity Quantification (SQ) are included in real-time energy feedback during design. New approaches using cloud computing, e.g., Naboni, Maccarini et al. (2013) and Naboni, Zhang et al. (2013), are a step towards using offloading intensive parametric runs to paid servers. This approach could also, naturally, be used in simulation with in-built sampling routines. As reader may see from the discussions in Chapters 1 and 5, interpreting the outputs from a simulation with multiple weather files is not as easy as the output of a single weather file.

Issues of complexity and speed have been looked at since the 1970s, when the drivers were the expense and paucity of computational power, and the challenge of introducing simulation to the architectural and building professions. While the original advantages of reducing simulation time by incorporating smaller weather files should now be irrelevant, the increasing complexity of building simulation codes has negated much of the gain in computational speed (Kershaw, Eames et al. 2011). These issues have given rise to several approaches based on simplified physical models and emulator-based approaches, which are discussed in Section 2.5. The approach described in Chapter 4 is also data- or emulator-based, though the synthetic weather series described in Chapter 3 may be used directly with a full BPS program.

2.5 Simplification and Emulation for Speed

The valuable capacity of the human mind to simplify a complex situation in a compact characterization becomes dangerous when not controlled in terms of definitely stated criteria . . . [because] the effectiveness of an argument is often contingent upon oversimplification ...

Simon Kuznets, *National Income, 1929-1932.*
Cited by Sidin Vadukut (2014).

Simplified methods are generally used to make computationally-expensive analyses tractable. In this capacity they can serve, in conjunction with statistical techniques, to explicitly calculate uncertainty, sensitivity, or risk. They could also be used in cases where doing a full building simulation is not worth the user's time and effort because, for example, not enough is known about the project to trust the simulation.

Zhao and Magoulès (2012) classify developments in simplified building simulation or prediction techniques into five categories: data-based techniques or "statistical" methods; simplified energy modelling techniques or "engineering methods"; neural networks; Support Vector Machines (SVMs); and "grey" methods (where some information about the system is available). In this thesis, we applied a technique called Gaussian Process regression, which is in the same category of techniques as SVM. Foucquier, Robert et al. (2013) also review methods in simulation, making a similar classification of methods into 'white', 'black', and 'grey'. The difference between these three categories is in the ability of the user to 'view' the underlying equations. The reader is directed to these two review papers for a review of approaches published so far.

We view this work from the perspective of computational efficiency for UA and SA studies. The issue is that the time required to simulate, say, representative samples of uncertain quantities or a full factorial experiment, can sometimes get out of hand. Emulators and simplified methods are both commonly used strategies in addressing this. In our work, we too found that regression is a useful approach. However, we are not convinced that classical regression models based on large pre-simulated sets are the way forward. Thus, in Chapter 4, we propose an approach that is able to build an emulator on-the-fly, and which can deliver explicit estimates of uncertainty using Gaussian Process regression.

2.5.1 Simplified Physical Models

Heat penetrates,
like gravity,
all the substance of the universe,
its rays occupy
all the parts of space.

Translated by Parag Rastogi,
23 May 2016

Le chaleur pénètre,
comme la gravité,
toutes les substances de l'univers,
ses rayons occupent
toutes les parties de l'espace.

Joseph Fourier,
Théorie analytique de la chaleur

The simplest description of BPS is that it solves a series of heat transfer problems. Of these, the central equation is the partial differential heat equation known as the Fourier equation

$$\frac{\partial^2}{\partial x^2}\theta(x,t) = \frac{1}{\alpha}\frac{\partial}{\partial t}\theta(x,t), \tag{2.1}$$

where $\theta(x, t)$ is temperature as a function of space (x) and time (t); and α is the thermal diffusivity of a material (Clarke 2001). Analytic approaches to solving this equation are based on some sort of domain transform, usually Laplace or Fourier. In other words, they seek to solve the differential equations of heat and mass transfer as algebraic equations. Haghighat and Athienitis (1988) compare the use of time and frequency domain programs. Boland (1997) propose the use of Laplace transforms and Duhamel's theorem. Other work using Fourier domain or discrete frequency transform-based solutions is presented and reviewed in Athienitis, Chandrashekar et al. (1985), Athienitis, Sullivan et al. (1986), Athienitis, Sullivan et al. (1987) and Athienitis and Santamouris (2002). Another popular approach is through the use of a variety of admittance functions, e.g., Clarke (2001) and McQuiston, Parker et al. (2005). See Sodha, Kaur et al. (1986) for a comparison of admittance and Fourier-based methods. See Clarke (2001, chpt. 2) for an example using yet another approach: the response factor method.

Donn, Selkowitz et al. (2012) propose an intriguing concept called the 'Building Performance Sketch'. Their idea is to provide building designers with a tool that allows them to evaluate the energy/comfort performance of 'sketches', i.e., plans with less details than would be required for a full-blown energy model. As the authors put it, "the performance analysis sketch is a model that can be created when no-one quite knows what the actual building will look like". According to the authors, the sheer amount of information required by Building Information Modelling (BIM) and Computer-Aided Design (CAD) tools hampers interoperability and experimentation. They mention recent efforts at translating between domain-specific 'views' of a building using high-level programmes or Graphical User Interfaces (GUI) that call separate modules, noting that this still requires a heavy investment of time on the part of the user managing the process. Schlueter and Thesseling (2009) developed an energy- and exergy-based analysis tool integrated with a commercial Building Information Modelling (BIM) software (as an add-on). The tool allows the designer to visualise different possibilities with real-time performance feedback. Although their tool still uses a large number of default values for parametric comparisons, it is a potential avenue for early-stage design exploration integrated with BIM.

Marsh and Carruthers (1995) created the first early-stage (simplified) design and analysis tool for environmental design (which eventually became the Autodesk Ecotect® software). They used object classes (backed by libraries of material properties) and interconnected specialised modular programs behind a graphical user interface. This allowed users to iterate designs without having to redraw the current design in each module for different kinds of analyses. While this approach is informative and saves users a considerable amount of time, it does not provide any “guidance”, leaving the user to understand the significance of the results.

Most of these approaches have since yielded to the major building simulation programs, since the original impetus for their development – computational efficiency – is not a major issue any more. The numerical solution of a thermal network is fairly efficient, and batch simulation permits the user to farm out jobs to servers, run simulations in the background, among other workarounds. Some of these techniques live on in the simulation programs as options, e.g., the use of the admittance method in EnergyPlus NREL and USDOE (2015).

2.5.2 Databases and Regression

If you torture the data enough, it will confess.

R. H. Coase,
cited by Gordon Tullock (2001)

Regression Analysis is a much (mis-)used technique for the modelling of relationships between quantities, or variables. In the context of BPS, regression techniques are applied to two situations: to complement full-scale simulation or to supplant it. Much of the work using regression-based emulators has been reviewed already in this chapter, in the context of UA and SA (Sections 2.4 and 2.7). Attempts to incorporate regression into the exploration of design options or simulation involving very abstract concepts are also common. We direct the reader to a forthcoming thesis from our colleague (Nault 2016) for a more thorough review of emulators.

Most approaches in this category focus on developing relationships between properties of interest to designers, e.g., insulation levels, and obtaining their relationships to final energy use. Examples of this include Amiri, Mottahedi et al. (2015), Asadi, Amiri et al. (2014), Coley and Kershaw (2010), de Wilde, Rafiq et al. (2008), Hygh, DeCarolis et al. (2012), Nault, Rastogi et al. (2015), Patidar, Jenkins et al. (2012), Perera, Halstensen et al. (2015), Tso and Yau (2007) and Wu and Sun (2012). Sensitivity analysis with *standardised regression coefficients* is another common use for regression, reviewed along with several other approaches in Tian (2013). See de Wit (2001) and Hopfe (2009) for discussions and reviews of the use of regression in uncertainty analysis.

A representative data-driven approach is that of Hygh, DeCarolis et al. (2012), who propose a parametric exploration of design choices based on regression analysis. They derive equations of relationships between factors and their outcomes from analysis of 20,000 simulation runs for each of four different climates. They found a strong linear fit between a composite equation of 51 factors/covariates (echoing Coley and Kershaw (2010), de Wilde, Rafiq et al. (2008) and Tian and de Wilde (2011a)) and the overall energy use for heating and cooling. The authors compared aggregated annual values for these parameters rather than hourly simulations. While this is a simplification, it is not unreasonable since annual values are handy metrics to guide energy-conscious design. In addition, due to the effect of internal heat gains and weather, hourly temperature profiles are sufficiently random and unpredictable that it is perhaps more sensible to compare broad trends anyway. While this method is hugely resource-intensive, it can conceivably be carried out for one's locations of interest to aid initial parameter exploration in those locations.

Burkhart, Heo et al. (2014) present a measurement and verification framework based on Gaussian Process (GP) models, showing that a GP fit using a Monte Carlo Expectation-Maximisation framework provides more robust predictions than one that does not consider uncertainty. They frame this in the context of the effort required to get high quality post-occupancy measurements from buildings. They propose that, with further extensions, their framework would be an efficient way to incorporate crudely estimated site data, whether uncorrelated or correlated, into emulators for building simulation. Work by Heo and Zavala (2012) also looks at GP models for measurement and verification. Yan, Kim et al. (2013) use GP emulators for optimal operation Wood, Eames et al. (2015) compared optimisation with a Genetic Algorithm (GA) and with a Gaussian Process regression-based emulator, and Kim, Ahn et al. (2013) also propose GP emulator-assisted optimisation. Eames, Wood et al. (2015) used emulators to simulate internal conditions due to climate change.

Ansuini, Giretti et al. (2012) proposed a probabilistic approach to decision-making in the conceptual design phase using "Probabilistic Design Spaces" composed of Bayesian Networks. The approach uses interconnected decision trees to model the process of design, where the relationship between a parameter (like material property) and its impact on an outcome (like indoor temperature) is assigned a probability based on data from a simulation or experiment. This is an empirical method, where the results of case studies are used to assign likelihood estimators to outcomes that cannot be easily represented by a closed-form equation. This approach relies on using data from case studies, which requires a significant investment in building those databases.

2.6 Weather input for Simulation

From the perspective of modelling and simulation, the "dynamic interaction between building systems and external climate is extremely complex, involving a large number of difficult-to-predict variables" (Guan 2009). Years of effort in this direction have yielded several useful rules of thumb and handbooks, and these, especially when they represent the synthesis of decades of painstaking research effort, form the bedrock of building design. However, unless a proper appreciation of the uncertainty inherent in characterising and predicting climate is included in design procedures (and measures taken to account for it), buildings could end up using more far more energy than planned for while periodically suffering overheating and other 'climate shocks'.

In practice, building simulation tools do not explicitly include uncertainty of weather inputs in a simulation run. In a typical building simulation work-flow, a single input weather file is chosen to represent the climatic conditions that a building would be expected to experience in its lifetime. In this thesis, for example, we use files the TMY files (Wilcox and Marion 2008), and those generated by METEONORM (MN) (Remund, Mueller et al. 2012a) for demonstration. We contend that this is an untenable state of affairs, and the stochastic nature of weather/climate inputs must be incorporated explicitly to gauge the sensitivity of a building to these inputs. Before we do that though, we present here the development and current practices in weather input for simulation. The nature of weather data used in building simulation has changed significantly over the years, from the 'typical weeks' proposed by Degelman (1997) to the most recent Typical Meteorological Year - Version 3 (TMY3) (Wilcox and Marion 2008). The list of data types and sources in Harriman, Colliver et al. (1999) has not changed much, except that TMY-like files are now available for thousands of locations worldwide.

This discussion is not about weather/climate data for the *design* of HVAC systems. The procedures for designing HVAC systems rely on *extreme* conditions, rather than typical ones, because the system has to be able to provide the necessary cooling, heating, or ventilation needs almost all of the time. For example, using the 99% design temperature implies that one's system will meet the cooling needs for 99% of the expected outdoor temperature values expected. Procedures for calculating these loads may be found in textbooks, e.g., Grondzik, Kwok et al. (2011) and McQuiston, Parker et al. (2005), or from professional manuals. Professional bodies like ASHRAE provide climate data for HVAC design (e.g., *ASHRAE Standard 169-2013*).

2.6.1 Typical or Reference Weather Years

The earliest efforts to create some sort of typical year data for simulation were the reference years from The Chartered Institution of Building Services Engineers (CIBSE) (Hitchin, Holmes et al. 1983; Holmes and Hitchin 1978), and those from the National Climatic Data Center (NCDC)[16]. Since then there have been several revisions and parallel efforts to define some sort of 'example year', 'test year', 'test reference year', 'design year', 'standard year', etc. These terms are equivalent, and we will use the specific terms from each of the publications reviewed below. In this review, we focus on algorithms of interest, instead of a chronological summary of development. The reader is referred to Crawley and Huang (1997) and Clarke (2001, ch. 7) for the historical development of reference or standard input data for building simulation.

Clarke (2001, ch. 7) defines a TRY as a "weather collection which, when judged against some relevant criteria, is deemed to be representative". Lund (1991, 1995) summarised the characteristics of a DRY file as part of a report for *IEA Solar Heating and Cooling Programme Task 9 (Solar Radiation and Pyranometry).* The basic requirement of a DRY file is that it should correspond to an *average year*, "both regarding monthly or seasonal mean values, and occurrence and persistence of warm, cold, sunny or overcast periods". The report reduces this to three fundamental requirements:

True frequencies Mean values should be as close as possible to the true mean, as obtained from long term measurements[17]. There should also be 'natural' daily patterns.

True sequences The duration and order of episodes must be representative of the long term prevailing climate.

True correlation The relationships between different meteorological parameters should be as accurate as possible.

The requirements we impose on the synthetic files proposed in this thesis are broadly derived from these[18]. The older Weather Year for Energy Calculation (WYEC) of Crow (1981) uses monthly means instead of Cumulative Density Functions (CDFs) (summarised in Crawley and Huang 1997; Gazela and Mathioulakis 2001).

[16] NCDC reference manuals *Test Reference Year* (TD-9706, 1976), *Typical Meteorological Year* (TD-9734, 1981) and Stamper (1977), cited in Clarke (2001, ch. 7)

[17] Which is still not the *true* population mean, but that is yet another question for our friendly neighbourhood philosopher.

[18] Tenacious readers will find the results and discussion in Section 3.10.

Ultimately, the quality of the resulting file is mostly determined by the quality of the input data. This is why the authors recommend that the DRYs should only be created for representative sites with high quality weather data available. The authors clearly state that a DRY should not be considered a “climatological description”. It is not intended for the sizing of HVAC systems either. It is meant to be used for predicting the expected average performance of buildings and solar energy systems. In general, the year-on-year variations in a climate are so great that the DRY generated for a fairly large meteorological area using just one high quality station would still be acceptably representative[19] (Lund 1991, 1995). Particular geographical conditions like relief and urbanisation must be taken into account, though, when choosing a station for calculating DRY.

As the report (Lund 1991, 1995) implies, the development of a typical year requires a significant investment in weather data gathering and processing. There are numerous algorithms proposed to process long term data into DRY files. Most of them are based on comparing frequency distributions, like the TMY algorithm (Wilcox and Marion 2008) described here and the one proposed by Festa and Ratto (1993). The procedures of Festa and Ratto (1993) and Wilcox and Marion (2008) are both based on the method proposed by Lund (1991) – they work with CDFs to identify a month that is ‘closest’ to the long-term probability distribution. One difference is that whereas Wilcox and Marion (2008) use the FS statistic, Festa and Ratto (1993) use a weighted average of the Kolmogorov-Smirnov statistic (KS statistic), the (absolute) difference between monthly standard deviations, and the (absolute) difference between monthly averages. Also, Festa and Ratto (ibidem) use *standardised variables*[20] instead of the raw variables. The literature is not clear on which algorithm is the ‘best’ for building energy simulations, because choosing a definition of ‘typical-ness’ is like choosing the ultimate climate classification: it will work some of the time, for some cases, but must be interpreted liberally to not become a liability.

2.6.1.1 Finkelstein-Schafer Years

There is no literal typical year type called this, but a clutch of very important years are based on the FS statistic: the TMY (Wilcox and Marion 2008) and its cousins based on the Sandia method, the CIBSE TRY year (Levermore and Parkinson 2006), etc.

Now in its third iteration, the Typical Meteorological Year (TMY) weather file developed

[19] That is to say, the temporal variation is generally big enough that a small amount of geographical variation is often unimportant.

[20] What we would call the z-scores.

by National Renewable Energy Laboratory (NREL)[21] is a widely accepted input for building energy modelling, despite its original intended use for solar energy applications(Wilcox and Marion 2008). The TMY algorithm uses four main variables for picking 'typical' months: maximum, minimum, and mean TDB; maximum, minimum, and mean Dew Point Temperature (TDP), maximum and mean wind velocity; and direct and global solar radiation (Direct Normal Irradiation (DNI) and Global Horizontal Irradiation (GHI)). These variables are disproportionally important in determining a building's energy performance and this makes a TMY-based simulation a fairly good indicator of a building's expected performance. The TRY algorithm from the UK (Levermore and Parkinson 2006) uses almost the same algorithm, just with a reduced set of only three inputs: mean TDB, mean GHI, and mean wind speed.

The procedure adopted for calculating 'typical' months is modified from an algorithm originally developed by Sandia National Laboratories (Hall, Prairie et al. 1978; Wilcox and Marion 2008, and reference 4 therein). Say the user has *N* years of recorded data. The months of each year are candidates for the typical year, so there are *N* candidate months for each typical month.

Steps 1 and 2 The algorithm begins by checking the closeness of each candidate month's CDF[22] to the long-term CDF of the recorded data, for the same month.This is done by calculating the FS statistic for each weather parameter of interest (Wilcox and Marion 2008, and reference 3 therein). The FS statistic is a metric that measures the 'average' deviation of a given CDF from the target CDF. A weighted sum of the FS statistics for the four parameters of interest is used to pick five candidate months (from the records) for each TMY month. The weighting is necessary to assign differing importance to the various meteorological parameters, and the scheme used in the latest TMY algorithm (v3) reflects its intended use in solar energy applications. Half the total points are for DNI and GHI, while the other half are distributed amongst the rest of the parameters.

Step 3 This step checks whether the candidate months had any persistent 'spells' of unusual mean dry bulb temperature and global horizontal radiation. For temperature, it is runs of consecutive days above the 67th percentile (warm spell) or below the 33rd (cold spell). For solar radiation, it is runs below the 33rd percentile (darker days). This persistence criteria excludes the candidate month with the longest run, most runs, and zero runs. Of the remaining candidate

[21] Golden (CO), USA

[22] That is to say, the empirical Cumulative Density Function (eCDF). In this thesis, CDF should be taken to mean the empirical Cumulative Density Function (eCDF), unless explicitly stated otherwise.

months, the highest ranked from the previous steps is selected as the TMY month.

Step 4 The final step is to concatenate the 12 distinct TMY months into a pseudo year. The discontinuities between the values at the boundaries of the concatenated months are smoothed for six hours on each side.

The International Weather for Energy Calculations (IWEC) and Canadian Weather for Energy Calculations (CWEC) data sets use the same procedure (Environment Canada 2015; Huang 2012). TMY files may be freely downloaded from the EnergyPlus website[23] or a third-party website[24].

2.6.1.2 ASHRAE Test Reference Years

The ASHRAE test reference year (TRY) uses a procedure based solely on dry bulb air temperature (Wong, Wan et al. 2012, and reference 17 therein). The procedure involves successively removing years from a long term record by using monthly mean temperatures. ASHRAE has a list of 24 priorities starting with 'Hottest July' and finishing with 'Hottest April'. The list alternates between hot and cold months (e.g., number 2 is 'coldest January'). If after removing years with each of these marked months, more than one candidate year still remains, the procedure moves through the list again by considering the next 'Hottest July', and so on (Stamper 1977). An intuitive criticism of this approach is that it may represent just an unreasonably 'mild' year, which may or may not have any relation to long-term averages. That is, a mild year in a record would always be selected as the TRY for that location, regardless of how different it is from the dominant climate of the area. Note that this procedure selects an entire year, instead of individual months like the other procedures. The original 'example years' proposed by CIBSE (Hitchin, Holmes et al. 1983; Holmes and Hitchin 1978) also selected entire years.

2.6.1.3 Extreme Years

The DSYs available for the UK are a "single contiguous year representing a hot but non-extreme summer" (Kershaw, Eames et al. 2010). Specifically, the third hottest summer in the reference data set, based on the external average summertime (April-September) Dry Bulb Temperature (TDB) (Levermore and Parkinson 2006). This is

[23] https://energyplus.net/weather
[24] http://climate.onebuilding.org/

a slightly different flavour of reference year from the rest since it is an actual year, like the ASHRAE TRY, rather than a year composed of months from different years. Levermore, Courtney et al. (2012) state that the standard DSY has "a 12.5% probability of being exceeded".

Crawley and Lawrie (2015b) propose an eXtreme Meteorological Year (XMY). The idea is to propose a set of files complementary to the TMYs, selected on the basis of highest daily/hourly maximum and minimum values of a set of weather parameters. These produce the daily/hourly minimum and maximum months. For example, the January with the highest average hourly temperature in the record would become the January of the hourly max TDB XMY. Instead of a weighted average of several weather parameters, like in the typical year algorithms, the authors tested separate XMYs for TDB, TDP, GHI, precipitation, and Relative Humidity (RH). Thus, 20 different files were tested. The hourly min/max XMYs based on TDB produced the largest variation, followed by the TDP- and precipitation-based years.

2.6.1.4 Typical Principal Component Years

Yang, Wan et al. (2011) have developed a different schema, called Typical Principal Component Year (TPCY), for generating typical weather data using Principal Component Analysis.Their schema is based on the finding that the climatic variance important to building simulation can be captured by considering just three variables: TDB, Wet Bulb Temperature (WBT[25]), and Global Solar Radiation (GSR, in MJ/m^2) (ibidem, and references 17 and 33 therein). They used the monthly averages of these three variables to construct a new monthly variable, called the Z-statistic[26]. They then selected those months from a 30-year record whose Z-statistic matched best with the 30-year long-term average Z-statistic[27].

This study found that the new TPCY gave similar results for building energy use when compared with TMY and long-term means. The TPCY performed marginally better than the TMY "in terms of the ability to follow the long-term monthly and annual building energy use estimation". The added value from using TPCY is not obvious from the results obtained from this paper for five cities in China. The authors go on to create TPCY for the next century using climate models, suggesting that the consideration of only 3 variables (compared to the 10 for TMY) could confer valuable

[25] We use Dew Point Temperature (TDP).

[26] The Z-statistic in their definition is the first PC, $Z = \alpha_1 \mathrm{DBT} + \alpha_2 \mathrm{WBT} + \alpha_3 \mathrm{GSR}$, where α_1, α_2, and α_3 are coefficients.

[27] Not to be confused with the *z-scores* used in this thesis, which are a standardised version of a single variable.

computational advantages in pulling data from climate models. This could be an advantage if hourly datasets for, say, the next century are available. Downloading and processing 3 x 876,000 data points is faster than 10 x 876,000 points, but the magnitude of this advantage may not be enough to justify abandoning the TMY algorithm, with its simpler and more intuitive mathematics.

The one aspect of the TPCY algorithm that the authors did not highlight is that it is not sensitive to the *a priori* weights assigned to different parameters in the TMY algorithm. The TPCY algorithm assigns weights to the different weather parameters based on their contribution to the variance when calculating each principal component. This could be yield better typical years for some climates, particularly those where solar radiation is less important than the other parameters in determining a building's energy consumption (since solar radiation has a disproportionately large weight in the TMY algorithm).

2.6.1.5 Others

A method proposed by Chan (2016) proposes using a genetic algorithm to determine the weights assigned to different meteorological quantities. They make a good point in that favouring representativeness of TDB and/or GHI only makes sense if one's system is sensitive to that. So instead of fixed weights, they use a genetic algorithm to determine the ideal weights on a set of meteorological parameters that minimise the FS statistic between the long-term CDF of each month and the CDF of the same month from each year. The method is computationally intensive, since it involves simulating the energy system/building with the entire recorded data set available (e.g., 35 years). In addition, the genetic algorithm operates on the energy simulation results from a set of twenty typical year files made by using a generation of candidate weights for the different parameters. It is difficult to see the generalisability of so much effort, given that the weights are specific to the simulation which has been used at each step. Chakraborty, Elzarka et al. (2016) demonstrate the use of SVM to forecast hourly solar radiation values over long time periods. Their method uses about 40-45 years for training, and was found to work well on 8-11 years of testing data.

There are many publications demonstrating the creation, or analysing the appropriateness, of different algorithms for typical year files in specific locations, mostly based on the TMY procedure described above. For example, Environment Canada (2015) created 'Engineered Weather Data' sets for Canada; Chow, Chan et al. (2006) reviewed typical years for Hong Kong and Macau, comparing TMYs and TRYs; Yang and Lu (2004) compared TMYs and Example Weather Years (EWYs) in Hong Kong; Oko and

Ogoloma (2011) created a TMY for Port Harcourt in Nigeria; Bulut (2010) created a typical solar year for the Aegean region of Turkey; Lund (2001) created typical years for Europe, Turkey, and Israel; Yang, Lam et al. (2007) examined the typical years for 60 cities in China; Kalogirou (2003) created TMY-2 files for Nicosia, Cyprus; Üner and İleri (2000) created typical weather data for 23 provinces in Turkey; and the German Weather Service (DWD) (2010) provides updated typical and extreme TRY for Germany which consider both recent climate change and urban heat island effects. Note that even the TMY, the ASHRAE TRY, and the CIBSE reference years were initially proposed for their respective jurisdictions (North America and Britain, respectively).

Degelman (1997) proposed that significant time savings could be achieved by simulating "typical weeks" for every month rather than the whole month. The speed gain they found was, firstly, not four times. Secondly, it was undercut by the need to reprogram the existing simulation software, which assumes contiguous data. We do not think this approach would help much with computation time even if the software was reprogrammed to accept only one week per month. The algorithms generally require a few days of "warm-up" before outputting valid simulation results. Hallgreen (1983) proposed a Short Reference Year (ShRY): "four climate sequences" of 14 days each. Once again, this is a useful short-cut to using the full 365 days, but less useful today. Interestingly, the authors used all synthetic data for these years. For example, temperature is generated additively using an annual mean term; a "yearly temperature variation" composed of four sine terms; a "daily mean temperature variation", generated by a second-order AR model with the noise term being the current cloud cover; and an "hourly temperature variation", which is modelled as a first order Moving Average (MA) process whose noise term is also the cloud cover, and which is damped by a sinusoidal term to account for the sun's passage. The cloud cover term is a Markov process.

Westphal and Lamberts (2004) present "simplified weather data" for Brazilian locations. This data consists of monthly averages for maximum/minimum TDB, atmospheric pressure, cloud cover, and RH; and two "typical days" to estimate heating and cooling loads. Predictably, the high mass cases did worse than low mass ones, since the influence of thermal inertia on load needs a larger warm-up period. Murdoch and Penman (1991) also presents a similar approach, by simulating a "reduced sample" of the annual weather.

The TMY approach was initially proposed for solar energy applications, but we did not include it in this grouping because it is probably the most widely-used algorithm to produce weather data for building simulation now. We discuss two interesting approaches to designed specifically for simulation of solar energy devices. Cebecauer and Suri (2015) proposed a method to generate typical years for solar energy applications, especially relevant for calculating risk in power production. They focus on reproducing months/years that represent statistically significant quantiles (e.g., 10th, 90th, etc.) of solar radiation quantities (GHI and DNI). Their procedure uses the CDF with the combined uncertainty from the model estimate and inter-annual variability to select specific months, like the TMY algorithm. The result is months/years that represent the various percentiles, or the chance of some aggregate solar radiation statistic exceeding a given threshold. Gazela and Mathioulakis (2001) propose a method based on minimising the "error in the monthly solar gain prediction" of a solar hot water system. This is similar to the proposal of Chan (2016) summarised below, in that it uses simulations from a system (in this case a solar hot water system instead of a building) to select typical months.

2.6.2 The Problem with Typical Data

All that is gold does not glitter ...

J. R. R. Tolkien,
The Lord of the Rings: Fellowship of the Ring

Clarke (2001, ch. 7) point out that most of the methods use "simple synoptic data"[28] to construct the typical years. The problem then, is that the user is not sure if the typical file has captured those aspects of the climate that are important to the system being studied. That is to say, *are we certain that all weather parameters matter equally to all buildings*?

Aguiar, Camelo et al. (1999) argue that "classic TMY are often built as general purpose tools", so may not be appropriate to the system being studied. We have no doubt that any one algorithm to select typical weather time series will fail to come out on top consistently for all systems in all regions. For example, Skeiker (2009) found the Sandia method to be the most representative for their case study using 10 years of data from Damascus, Syria, and a "typical Syrian building's thermal system". Argiriou, Lykoudis et al. (1999) found the Festa-Ratto method most appropriate for solar energy applications in Athens, Greece. 'Simulation of building energy and indoor environmental quality - some weather data issues' (1999) points out some of these issues,

[28] "Pertaining to or forming a synopsis; furnishing a general view of some subject; spec. depicting or dealing with weather conditions over a large area at the same point in time." (Oxford English Dictionary 2016)

suggesting that if the priorities for each building are different, then the ‘typical year’ for each building is different. Another recommendation is to create a three-year file: typical/average, cold/cloudy, and hot/sunny. Some authors, like Chan (2016) and Gazela and Mathioulakis (2001), address this issue by customising the weights for each system. However, the findings of Su, Huang et al. (2009) indicate that the weights may not be as much of a problem as previously thought. In an investigation of 3600 typical years for Beijing and New York, they found that the typical months selected using different weights for the key weather parameters[29] did not vary much. They also found a high probability of “significant overlap” between typical months selected using the “TMY/TMY2 weighting” and an ensemble selected from variable weights. The weights for each weather parameter were varied considerably, e.g., 0.06-0.5 for TDB. The use of synthetic files addresses this issue because there are no weights to be considered. Through simulation with the entire set of synthetic files, one gets a picture of the range of responses to be expected from any system. The method is computationally expensive, but this can be partially addressed with the use of an emulator like the one we propose in Chapter 4 during an exploratory phase, when the possible number of simulations is truly intractable. The issue that underpinned the development of typical data – computational time – should not be an issue now. We have come a long way in computing power since Hui and Cheung (1997) asked: ‘Multi-year (MY) building simulation: is it useful and practical?’

Using a single typical or reference file is conceptually and computationally far simpler than working with several files from a random generator, each of which have some probability of occurring. However, using typical files, of any sort, cannot be used to assess risk. Using TMY files in building simulation for “what-if” analyses tells us only about the response of a building to typical conditions, which are representative only for the period of record of the file. Kershaw, Eames et al. (2010) further argue that “there is a tendency for the external statistic to be confused with the internal”. Simulating the third hottest month (selected using the monthly mean temperatures) does not mean that the full range of weather conditions that will produce the third hottest month *inside* the building are present. In addition, using reference files does not allow the calculation of complex questions related to risk. They tested the results from using TRY and DSY files against long-term measured data used to create those reference years, from Plymouth and Edinburgh. They found that while the “... TRY allows rapid thermal modelling of building designs it is not always representative of the average energy use (compared to the average of the [reference data set]) and gives no indication of the expected range of energy usage for a particular form of architecture”

[29] TDB, TDP, wind speed, and GHI.

(ibidem). The DSY also "... consistently underestimates the levels of overheating and thermal discomfort within the building" (ibidem). Aguiar, Camelo et al. (1999) observed that the misrepresentation of climate by a reference year is exaggerated by the tendency of even "small variations in the mean level of the meteorological forcing ... to reflect non-linearly on the thermal behaviour of the simulated cells". They found that the "classic" TMY for Lisbon overestimated discomfort below 18°C by 19-38%, while underestimating discomfort above 24°C by 6-20%, when compared to recorded data. However,they found that both the "classic" TMY and "stochastic" TMY[30] "reasonably substitute for the real long term weather" (ibidem).

A typical year can not, or should not, be constructed from a short period of record. Using TMY files in building simulation for "what-if" analyses tells us only about the response of a building to typical conditions, which are representative only for the period of record of the file. Several studies have pointed out the sensitivity of simulation output to weather data, including but not limited to Bhandari, Shrestha et al. (2012), Chinazzo (2014), Crawley and Huang (1997), Hong, Chang et al. (2013) and Kershaw, Eames et al. (2010). The studies usually propose using measured weather data from the vicinity of the building to best characterise the climate *in all its states*, as it were.

It has been previously proposed (e.g., Chinazzo 2014) that a 'range' of possible performance outcomes, i.e., results from simulation runs with different weather inputs, is a better characterisation of the *range* of performance that a building will inevitably give. Simply summarised, the argument is that if one does not know exactly what (weather) inputs one's (building) system will experience, then one is better off knowing the effect of a range of possible inputs on it. A given weather file is, after all, a representation of one scenario out of an immense number of possibilities. Therefore, by using only one weather file, we are restricting ourselves unnecessarily to one 'experimental result'. If a building is never going to experience a narrow set of weather conditions exactly, i.e., the ones contained in a typical year file, then the quality, 'averageness', or ability to represent best the most typical weather, of said weather file is irrelevant. This is the reasoning behind the sensitivity analyses recommended in this thesis.

2.6.3 Weather Generators and Synthetic Data

We summarise proposals from various authors, mostly intended directly for building simulation. Weather generators for climate change inputs are in Section 2.2. The

[30] A TMY chosen from synthetic data generated for this station.

weather generators discussed in this section do not explicitly include climate change forecasts, with the exception of the MN software, which offers the option of creating 'future' typical files. For a survey of weather generators, see Richardson (1981). The IPCC (2013) classifies weather generator into two types: the rainfall-based generators ("Richardson-type"), or the serial generators. The Richardson-type generators model wet or dry days as a Markov process, and then use another distribution to predict the *amount* of rainfall (e.g., a gamma distribution). The UKCP09 generator is a Richardson-type generator, though the authors argue that "it is now widely recognised... that the clustered nature of rainfall occurrence is better modelled by more complex clustered point process rainfall models"(Jones, Harpham et al. 2010). The current model used is the Neyman-Scott Rectangular Processes model. The UKCP09 generator yields 3000 files for a location, until the 2080s. See Levermore, Courtney et al. (2012) for a discussion of the anomalies and subsequent correction in sunshine predictions from the generator.

A criticism of the Richardson-type approach is its inability to model long spells of persistent wet or dry weather. That, presumably, would be a serious issue in those parts of the world where there isn't a constant threat of getting one's head wet all throughout the year. For example, sub-tropical climates with clear wet and dry seasons. So, the alternative type, serial generators, models a sequence of wet or dry days. The serial generator presented in Racsko, Szeidl et al. (1991), for example, describes the length of a wet series as a geometric distribution. The length of dry series is modelled a mixture of two geometric distributions, with "probability $1 - p$ for the short series and probability p for the long series (longer then eight days)". The parameters λ of these geometric distributions are approximated as Fourier series with period 365 days. See Jones, Harpham et al. (2010) for a fuller description of the development of a weather generator. Normally, generators are calibrated using historical data, like 1961-1990 in the case of UKCP09. Weather generators may be useful in any of the following situations IPCC (2013) :

- records are too short for the task at hand,
- data availability is sparse,
- gridded data is needed for a spatial analysis,
- one wishes to check the effect of both mean climate and fine-grained variability.

METEONORM Handbook Part II : Theory gives an overview of the algorithms used to generate solar, temperature, and other weather parameters' profiles in the MN software. Many of the models come from the SoDa project[31]. The solar radiation generation starts from monthly values, and proceeds based on a modified version of the model originally proposed by Aguiar, Collares-Pereira and Conde (1988). The procedure uses Markov Transition Matrices (MTM) to "calculate a daily sequence of clear sky daily clearness indices (KT_d)"[32] (Remund, Mueller et al. 2012b). It provides an upper bound to the value of solar radiation at a location, since it is the amount of solar radiation that would be received on the surface of the earth when the sky is perfectly clear. The MTMs are hard-coded into the software, and were calculated with data from 121 stations. The relevant MTM is selected based on the monthly mean clearness indices (KT_m) and determines the state of the clearness every day. The hourly values are calculated using the TAG (Time dependent, Autoregressive, Gaussian) model of Aguiar and Collares-Pereira (1992). The model works by separating the hourly clearness index prediction in to an "hourly clearness index of the average daily profile" and a first-order AR function (Remund, Mueller et al. 2012b). The software also offers the ability to generate minute-to-minute radiation values.

The generation of temperature values is "based on the assumption that the amplitude of the temperature variation during daytime is approximately proportional to the amplitude of the daily global radiation profile" (ibidem). We made a similar observation in our work, and used the relationship to generate hourly solar radiation profiles. The generation of the daily temperature profile is based on a factor kx, "the ground to extra-terrestrial irradiation ratio". It is a ratio of the cumulative "radiation received on the ground since sunrise, to the amount of solar radiation that a surface perpendicular to the... [sun's rays]... would have received during the same period". The temperature is positively correlated to kx. The temperature profiles are influenced by the clearness of the sky, characterised by a "nebulosity index", originally proposed by Perraudeau (referenced in ibidem). The statistics of the generated data were checked against measured data from 10 stations, and were found to be well-reproduced. The procedure, however, does not reproduce the variance and extremes seen in the recorded data. The software generates a host of other parameters (e.g., cloud cover), though these were not validated to the same extent as solar radiation and temperature. The procedures in MN check that the statistics of the results of each random number simulation (e.g., the AR models), match those of the recorded data, where available. The MN software also does geographical interpolation, since the project was only able to use recorded data from about 5000 stations across the world. The jury is out on whether geographical interpolation is better than using files from nearby stations, though the

[31] http://www.soda-pro.com

[32] The clearness index, usually K_t, is the "ratio of the total radiation on a horizontal surface to the extraterrestrial global solar radiation on a horizontal surface at the same time" (Solar Energy Laboratory 2009).

authors' self-checks found the "errors are mostly within variations of climate from one year to the next" (Remund, Mueller et al. 2012b).

Degelman (1976, 1991) proposed a weather data generator for a whole host of weather variables. It relies on hard-coded "long-term statistical averages and standard deviations", and a model with a "deterministic portion" and a "probabilistic portion". The deterministic portions are the "general shape of a dry-bulb temperature [TDB] curve" and earth-sun geometrical variables. The probabilistic portions include a cloudiness model based on persistence; a TDB generator which daily maximum, average, and minimum temperature; a daily dew point depression (for TDP); pressure curves adjusted experimentally to mean periods of about 3 days; and wind speed that is roughly negatively correlated with solar radiation. Effort was made to preserve the influence of solar radiation on TDB. See Degelman (2003) for details and worked examples of these concepts.

Adelard, Boyer et al. (2000) and Adelard, Mara et al. (1999) describe the development of a synthetic data generator – RUNEOLE – that is able to generate synthetic sequences of interest on demand. The work is focussed on the island of Reunion, in the Indian Ocean. The authors use a library of empirical models from several authors (Table 3 in Adelard, Boyer et al. (2000)) and stochastic models. The software picks the appropriate set of models based on climatic quantity and region. They account for the cross-correlations using neural network models. All of this is implemented in a GUI, where the user is able to select the variables and days they wish to obtain weather data for. After those initial papers, the next mention of this generator that we found was in David, Adelard, Garde et al. (2005) and David, Adelard, Lauret et al. (2010), where the authors have used RUNEOLE to generate typical weather data for simulation. Incidentally, these authors find that the "autocorrelation of each climatic variable may be described by a first order linear autoregressive model". In Chapter 3, we show that our example climates have significant auto-correlation in TDB residuals up to 3 or 4 lags. Then again, we did not include a tropical maritime climate like Reunion in our database.

The weather data generator (Type 54) included in TRNSYS (Solar Energy Laboratory 2009) also begins with monthly average radiation, humidity, and temperature. The radiation model is based on the familiar clearness index K_t. There is a choice of two temperature models: a stochastic one based on a second-order AR process, or a deterministic one. The coefficients of the AR model are hard-coded, based on data from three stations in the US. The AR process models the "hourly deviations" from the monthly average of the TDB. The deterministic model interpolates a cosine between the minimum and maximum daily temperatures. In this case, the daily average and maximum values "are obtained from normal distributions where the means and standard deviations are either Input or estimated from correlations" (ibidem). In comparing the two, the authors leave the choice to the user, since "the stochastic model ... better represents the hourly autocorrelation structure of the dry bulb temperatures ...[but] does not always generate temperature data with the correct daily autocorrelation and daily distribution. The deterministic model ... consistently reproduces the daily structure but neglects the variation and autocorrelation of the hourly sequence" (ibidem). The humidity model is a model of the "dew-point depressions". The daily average TDP is sampled from a normal distribution "with the mean equal to the monthly-average dewpoint temperature and the standard deviation equal to the standard deviation of the daily maximum dry bulb temperature" (ibidem). The biggest shortcoming of this generator is that it *does not preserve cross-correlation.* The methods are based on refs. 1-6, 10-12 in the manual.

The work of Scartezzini, Bottazzi et al. (1987, 1989) proposes the use of two 'stochastic' weather methods: simplified stochastic simulation (SSS) or Repetitive Meteorological Day (RMD). In both methods, the user simulates a certain number of day profiles per month (say, 4) and then uses their historical probability of occurrence to 'assign' each of these profiles to the month, creating a month-long simulation. The difference lies in how they generate their typical days. In Scartezzini, Nygard Ferguson et al. (1990), the authors propose Markov and Auto-Regressive Moving Average (ARMA) processes to model hourly GHI and TDB. They use 10 years of data to calculate the parameters. Solar radiation is modelled using two Markov chains – one for 'daily insolation ratio' (daily clearness index) and one for 'hourly atmospheric transmittance' (hourly clearness index) – and five probability matrices to characterise the transition between different types of day and hours. The hourly temperature profile is additively composed of a daily profile (the daily 'shape' of Degelman (1976)) , a slope, and a residual. The temperature profile depends on the 'type of day', through 'average temperature profiles' for specific days calculated from historical data. A 'slope' of daily temperature profiles is selected for every day from a Gaussian distribution. Finally, the residual is a first-order AR process. Various shifting operations keep the day-to-day profiles aligned.

One question that arises in the creation of any synthetic data is its advantages over recorded data. If long-term, high-quality, data is available for some location, is there any point in using synthetic data? As Kershaw, Eames et al. (2011) point out, the utility of recent records in predicting future return periods (i.e., probabilities of weather events of interest) is limited by the length of the record. For example, if a 100-year event (i.e., over a long enough record, this event will occur roughly 1% of the time) happened thrice in the last 10 years, that does not make it a 3-year event. While the return period obtained from any weather generator is speculative, it does at least provide bounds on a system's response. Then, it is the decision-makers who must choose the probability for which they would like to design. For example, HVAC system failure may be acceptable after some value of outdoor temperature or episode of some intensity, which has a very low probability of occurrence. Kershaw, Eames et al. (2011) warn that using the UKCP09 weather generator to assign return periods should be done with "extreme care", and "... return periods longer than 5-years should be used with caution".

Hong and Jiang (1995) and Lee, Sun, Hu et al. (2012) proposed a variance process for creating stochastic weather variables. This approach is very similar to ours, except that our models are based on temperature, they work with residuals, and we consider both AR and MA effects. A variance process is theoretically cleaner, since one is able to code in both auto- and cross-correlations explicitly. However, we found them cumbersome to work with and difficult to train reliably. Note that Hong and Jiang (1995) use the variance for daily values, and the 'shape' approach of Degelman (1976) for hourly values. The approach of Lee, Sun, Hu et al. (2012) involves transforming the data, and final recalibration of the CDFs to match the training data. van Paassen and Luo (2002) also propose a weather generator for future climate based on AR relationships. In their application, for The Netherlands, the driving parameter is the "type of day", one of 11 types ranging from very cloudy to bright.

The approaches discussed here all generate high-resolution data from low-resolution historical records. This puts them in the same category as several proposals, like those of Aguiar, Collares-Pereira and Conde (1988), Boland (1984, 1995), Hansen and Driscoll (1977), Magnano, Boland et al. (2008) and Magnano (2007), whose results have been useful to our work. Our approach builds on the results of these studies by for example, using Fourier series. However, we are working in a different framework, one where computational burdens are less relevant, typical data is widely available (often using these authors' work), and the goal is the demarcations of confidence intervals for uncertainty and sensitivity analysis. See de Livera, Hyndman et al. (2011) for a general overview of the analysis of time series with seasonal components.

PRÉCIS

- The interaction of weather and building properties is complex and non-linear.
- Weather cannot be parametrised for conventional sensitivity analyses like those with factorial experiments.
- The sensitivity of simulation results may only be examined via a Monte Carlo simulation, using synthetic weather. This would also establish empirical confidence bounds on the outcomes.

2.7 Uncertainty in and due to Weather Inputs

The difficulty of fully characterising a system that depends on the climate is that we cannot fully characterise the climate itself, especially future climate.

Fürbringer and Roulet (1995) state that building simulation needs a sensitivity (or uncertainty) analysis because the input "data used in simulation have large confidence intervals"[33]. We have already discussed that future weather data has large, irreducible, uncertainties. As we have summarised above, the bulk of uncertainty and sensitivity studies have focused on analysing the effect of uncertainty in material inputs, model assumptions, and variations due to occupant behaviour. Uncertainties in the weather input arise due to several factors, including:

Modelling assumptions like simplifications of physical phenomena, and skipping phenomena that are not well understood;
Incomplete records used to feed and calibrate climate models;
'Downscaling' of global circulation models to a region of interest, which creates biases and inhomogeneities;
Microclimate effects due to the intervening natural and built environment and topography between a building and the weather station used to represent its location.

Arguably, the first attempt to formally analyse the sensitivity of buildings to climate with simulations could be the Climate Severity Index (CSI) developed by Clarke (2001), Markus, Clarke et al. (1984) and Markus (1982). Of course, the idea of controlling sensitivity to climate is nothing new. The uncertainty introduced by using different statistical methods to obtain average, typical, or representative data for a location is the focus of this section. Some studies summarised here also try to quantify the

[33] Technically, the authors were talking about variability intervals.

effects that arise from the difference in precision between different parameters in the same dataset. Some weather parameters in these datasets are less precise than others since they are not directly measured, but instead modelled from other parameters measured on site, e.g., solar radiation. We divide the uncertainty in weather/climate inputs into two categories: spatial and temporal. As must be clear to the reader, it is the latter that is the focus of this thesis. After that, we will discuss various studies that have examined the different sources of variation in weather inputs.

2.7.1 Temporal Uncertainty: Climatic Volatility and Climate Change

The complexity and uncertainty of climate data seriously hampers the reliability of predictive simulation. The inherent year-on-year unpredictability of meteorological parameters are not captured by long term means of that parameter (a commonly reported and recorded statistic). Temporal uncertainty has traditionally been the *raison de l'emploi* for whole castes of soothsayers, diviners, and more recently, meteorologists and statisticians. It arises because deterministic forecasts are precise but inaccurate beyond a few days ahead, while stochastic forecasts could be accurate but are deliberately imprecise. The standard approach to temporal uncertainty in building simulation is to ignore it with the use of a historical or synthetic typical weather file. This is acceptable if the climate is stable and the quality of the typical file is assured. As we have discussed earlier, neither of these are tenable assumptions. A major component of temporal uncertainty is climate change, or the lack of climatic stability. Naturally, the uncertainty in climate predictions is directly proportional to the distance from the present. This is not strongly reflected in our work since the year-on-year uncertainty is far greater than the slight change in prediction uncertainty over the years. Nevertheless, it is prudent to weigh far-future predictions less than near-future ones. We reviewed some of the literature addressing temporal uncertainty in the discussion of weather inputs in Section 2.6.

There are few studies which explicitly include uncertainty by working with probabilities rather than fixed points. As de Wilde, Rafiq et al. (2008) point out, "...the majority of R&D efforts in building [are] still performing deterministic rather than probabilistic work, working with fixed input parameter values across the board and representing results as single values rather than probability ranges. Yet such approaches do not seem sufficient when dealing with long-term prediction of the thermal behaviour of buildings under climate change". They found that the interaction of uncertainties in building construction and operation, while large by itself, was further magnified by uncertainties in the climate. de Wilde and Tian (2010) examined the performance metrics, modelling assumptions, key design parameters, and zone resolution of studying the impact of climate change on an office building in the UK. They found a slight decrease in annual carbon emissions, uneven overheating risk, and a small effect of zoning resolution on results. Tian and de Wilde (2011b) examined the propagation of uncertainties in climate predictions. They found that the "uncertainty in predicted annual cooling energy is significantly higher than the uncertainty in predicted heating energy", and that the uncertainty due to climate becomes increasingly dominant from the present to the 2050s. Coley and Kershaw (2010) found a linear relationship between changes in mean and max external and internal temperatures. The strongly linear character of the relationships is indeed striking. The authors term the slopes of the lines *climate change amplification coefficients*. If this is the case for most buildings, then it would make the investigation of uncertainty far easier. While we did not use the exact same test, we do agree that the response of buildings to changes in aggregate weather quantities tends to be *smooth*.

One style of speculative studies involves the use of weather files from different cities to model climate change. That is to say, speculating the effect on buildings if the climate of, say, Geneva becomes like Rome, Italy, or Nice, France. We have published one such study ourselves, in which we examined the effects of climate 'shifting' on dwellings located in three climates – two in Europe and one in India (Rastogi, Horn et al. 2013). We picked nearby locations based on temperature contours and tested the effect on overheating hours and enthalpic distance to a fixed comfort zone. We noticed no strong trends for comfortable hours in any of the locations/buildings (figs. 5-7 in publication), though the Alpine European location did show a discernible linear trend with average annual outdoor temperature difference. The graphs we presented (figs. 2, 5-7) provide an easy way to estimate the climatic robustness of a building. If a building's performance changes sharply when moving away from its 'home' climate, the building is sensitive to outdoor conditions, and vice-versa. Using only temperature, however, gives uncertain results since temperature alone does not determine the climate. The work of Eames, Wood et al. (2015) and Gaterell and McEvoy (2005) takes the same approach. An advantage of this type of study is that the "future" weather data is real, so it preserves correlations and other hard-to-reproduce features. On the other hand, topography and relief play such an important role in climate that it is hard to imagine the wholesale (neat) shifting of climates northward (southward in the southern hemisphere).

Among non-climate change studies, most compare the simulation of typical or design data with actual years. Hong, Chang et al. (2013) found that TMY files for a variety of climates over- or under-estimated building Energy Use Intensities (EUIs) by 4-6%, HVAC EUI by approx. 5-10%, and peak electric demand by 5-20%. More worryingly, the predicted percentage reductions in peak electricity demand from energy efficiency measures was off by 10-30%. These are a little less than the figures we have found in studies with collaborators (Chinazzo 2014; Chinazzo, Rastogi et al. 2015a,b). Some of those results are discussed in Chapter 3. The aforementioned studies usually propose using measured weather data from the vicinity of the building to best characterise the climate.

A strain of research, usually sponsored by standard-setting bodies like ASHRAE and CIBSE, has looked into the length and completeness of records that are necessary to assess the climate of a location. Vignola and McDaniels (1993) state that 5 years of data is enough for estimating long-term monthly average solar radiation, while 15 years demonstrates the "variations experienced from year to year". In addition, they found that it would be possible to discern changes in climate to a high level of statistical confidence from a 30-year data set. Hubbard, Kunkel et al. (2005) analysed a host of issues in weather data collection to extract "minimum criteria with respect to record length and completeness that will allow reliable estimation of the Design Weather Conditions (DWC)". Using 33 years of data from 17 special 'baseline' stations (see their paper for selection criteria), they compared the effect that different inhomogeneities (such as instrument problems, measurement errors, and outliers and trends owing to the observation system) in the weather data have on the accuracy of the DWC values. The authors found, for example, that the most significant amount of error is seen in psychrometric measurements, usually due to measurement errors in the hygrometers themselves and in the associated dry bulb temperature sensors. The authors conclude with the recommendation that at least ten years of data is required from a station in order to include it in the *ASHRAE Handbook: Fundamentals*. While this is not an explicit discussion of *uncertainty*, it is an effort to quantify what constitutes a *reliable* weather input.

2.7.2 Spatial Uncertainty: Natural and Human Factors

We have by and large ignored this aspect of weather uncertainty in this thesis, mostly because it is a very different sort of problem from temporal uncertainty. At its most basic, spatial uncertainty can be said to arise because we do not have weather records for every conceivable building site in the world. Moving away from a weather station entails a loss of certainty because geography gets in the way: terrain, changes in elevation, the presence of water bodies, etc. The point is then, that even having long high-quality records for a location means that the idiosyncratic weather of a particular site cannot be well accounted for. This problem is most acute in regions with sharp changes of relief, like Chile, or those with sparse station coverage, like parts of Africa. All of this before we even consider human factors. Some software and consultants offer geographically interpolated data, e.g., Remund, Mueller et al. (2012b).

Eames, Kershaw et al. (2012b) examined the appropriateness of a 5x5 km grid for generating future climate files based on the UKCP09 projections. They found that the "spatial variability of the weather is the dominating factor", i.e., the "current spatial variation... is greater than the predicted future climate change". While it is not particularly surprising that the variation in climate over about 12 degrees of latitude and 7 degrees of longitude (the rough geographical extent of the UK) is more than that predicted by climate change, this position could be reversed in the future. The authors find that the predictions "exaggerate the differences in temperatures between the north and the south of the UK". Echoing the concepts of Kriging, they recommend that spatial interpolation cannot be applied uniformly across the UK, because the differences between adjacent grid cells are not uniform. The paper ends on a cautionary note by saying that it is "prudent" to use the high resolution future grid, to within the limits of computational capacity available.

According to Bhandari, Shrestha et al. (2012), there are four methods used by vendors to provide weather data sets for locations not sufficiently close to a measurement station: satellite data with a geospatial resolution of 15-40 km^2; measured data from the nearest weather station; statistical interpolation from several nearby stations; or, a gridded Mesoscale climate simulation model seeded by nearby sensor data. They tested weather data for their Oak Ridge[34] campus obtained from commercial sources with a weather station on site. Their purpose was to compare the discrepancy in results from energy simulations that may arise from differences in how weather data is interpolated for sites that are not sufficiently close to a weather station. They found that peak differences in weather data from different sources are higher for daily or hourly data than monthly averages. This makes intuitive sense, since it is expected that the systematic error introduced by choosing one interpolation methodology over another would cancel out over a month, but still produce large discrepancies for individual data points and short averaging periods. This study found that different weather parameters had different levels of discrepancy between measured data and individual data sets. We found a similar result in our examination of typical weather files from two different sources – the Energy Plus website and MN software – and nearby stations, e.g., the airports around Milan, Italy (Chinazzo 2014; Chinazzo, Rastogi et al. 2015a,b). For energy usage, the authors found that whereas overall energy consumption shows

[34] Tennessee, USA

a discrepancy of only 7%, “the heating and cooling loads differ by ± 40%”. The authors used TMY3 data from a nearby airport for the energy comparison. Upon replacing one weather parameter at a time in the TMY data set with one of the commercial data sets, no clear pattern emerged, except that changes in each parameter influence the simulation results differently.

Considering human factors, a major topic of concern, and research, is the effect of urbanisation. Broadly described as the ‘urban heat island’ or urban micro-climate, it has proved to be fiendishly complicated and difficult to model. It is by and large expected to exacerbate the effects of climate change. As the name suggests, urban heat islands entail an elevation of temperature, reduction in wind speed, and increase of pollutant concentration. In heat waves or even regular hot weather, the conditions in cities is much worse without the possibility of ventilation, and with the oppressive, polluted inner city air. In cities with widespread air conditioning, the proximity of several heat pumps working to dump ever more heat into restricted urban canyons is not a pleasant prospect. By and large, urban infrastructure has been found to act like thermal storage for the urban micro-climate. McKitrick and Michaels (2007) call attention to the direct impact of urbanisation on weather stations in their discussion of “local land surface modifications and variation in data quality”. These include urbanisation and other changes in land use around a weather station, the general socio-economic development of the area surrounding a measuring station, among others.

The work of Crawley (2008) and Levermore, Courtney et al. (2012) are two examples of perhaps a handful of studies that have formally looked at the combined effect of both temporal and spatial uncertainty in weather inputs, for simulating future cities. The UK case studies often cite the amplification effect of urban heat islands, e.g., Ren, Shankland et al. (2012). We have recently completed, with colleagues, a qualitative study into the interaction of urban and climate change effects on the uncertainty of predicting future overheating for inner-city residences in Geneva (Agarwal, Rastogi et al. 2016). Another co-authored publication looks into the change of indoor conditions in a historical building due to extreme urban transformation in Sao Paolo, Brazil (Pastore, Rastogi et al. 2016). Most authors do, however, show a familiarity with the broad outlines of the interaction between urban micro-climate and climate change.

2.8 Summary

In this chapter, we have discussed the state of the art in the use of building performance simulation, especially as it relates to the analysis and quantification of uncertainty and sensitivity. The main argument presented in this review of literature is summarised below. These first two chapters focussed on two arguments: first, for incorporating Uncertainty Analysis (UA) and Sensitivity Analysis (SA) into the building simulation workflow, i.e., working in a stochastic paradigm; and second, the issues in the procedures and algorithms currently available for this. These issues include generalisability of algorithms/procedures, ease-of-use, complexity, and computational load. We will return to these issues throughout the thesis. Several choices in subsequent chapters were motivated by the need to address the issues raised in this chapter including, but not limited to, using only typical year files, using probabilistic regression fits, among others.

PRÉCIS

- Energy-conscious building design is desirable, though with an acknowledgement of the variety and vagueness of thermal criteria for comfort.
- Building simulation permits the exploration of what-if scenarios in the design phase.
- The climate is a dominant boundary condition for the thermal performance of buildings, in conjunction with user behaviour.
- The uncertainty due to climate can be reduced *to some extent* with better information, but it cannot be eliminated. Especially uncertainty about future weather.

■ Energy-conscious building design could benefit significantly from uncertainty and sensitivity quantification of the weather/climate input.

3 Synthetic Weather Inputs for Building Simulation

The Moving Finger writes; and, having writ,
Moves on: nor all your Piety nor Wit
Shall lure it back to cancel half a Line,
Nor all your Tears wash out a Word of it.

Omar Khayyam (ca. 1048-1122),
The Rubaiyat of Omar Khayyam, ruba'i 71.
[translator F. Scott Fitzgerald, Fifth Ed. (1889)]

3.1 General Approach

So far in this thesis (Chapters 1 and 2), we have built up a case for why random weather inputs should be used in simulation to account for an uncertain climate. An energy simulation that accounts for uncertainty in the weather input, or alternatively, assesses the sensitivity of a design to weather variation, may lead to more robust design solutions than a process that uses deterministic weather inputs only. A method to generate these random weather inputs efficiently and simply is presented in this chapter. Before we expound the details of the procedures, we would like to remind the reader that these synthetic weather time series are not *predictions*. They are meant to be understood as tools for what-if analyses. The inclusion of the climate change forecasts is an extension of the analysis to include probable future mean changes in climate.

We only create synthetic generators for three time series: Dry Bulb Temperature

(TDB), Global Horizontal Irradiation (GHI), and Relative Humidity (RH). A further four time series are affected: Direct Normal Irradiation (DNI), Diffuse Horizontal Irradiation (DHI), Humidity Ratio (W), and Dew Point Temperature (TDP). Future work could extend this approach to other parameters like cloud cover. However, the cross-correlation or dependence structures become progressively more complicated as more and more time series are considered, and several time series will need to be modelled as non-normal variables. After reviewing the relevant literature (Section 2.6.3), we find that our generator is unique in two respects: we use a very short weather record (a typical year), and the primary generator series is Dry Bulb Temperature (TDB). In addition, an important finding in this thesis is that the *underlying structure of the models used is the same in all the tested climates*, which has not been demonstrated before. The fitting of robust models to only one year of data has also not been demonstrated before.

We demonstrate our method with Typical Meteorological Year (TMY) files from Geneva, Switzerland; New York, USA; and Delhi, India (details in Table A.1). New York has three stations – John F Kennedy Airport (JFK), LaGuardia Airport (LAG), and Central Park (CPR) – while the other two have one station each at their respective airports. In addition to the plots presented in this chapter, results are also presented in Appendix A. Due to time and space limitations, we limited ourselves to exploring climate change scenarios for Europe (Geneva). Results from sample simulations[1] using the generated synthetic time series are presented in Section 3.11. We begin by describing the 'splitting' procedure, separating the weather time series into random and deterministic components. Then, we describe the simulation of the residuals to produce variations on the source time series, and examine the results. Additional concepts and discussion, especially for time series models, are in Appendix A.

3.2 The Synthetic Generation Procedure

The schematic presented in Figure 3.1 details the procedure for generating synthetic weather files. Later, Figure 3.5 shows a variation that incorporates forecasts of daily values using different climate models and Representative Concentration Pathways[2]. Details about each step of the procedure are in the rest of the chapter, and some key underlying concepts are discussed in Appendix A. The symbols used in Figure 3.1 are used consistently throughout the thesis.

The most important restrictions placed on the synthetic weather data generator is to

[1] In this thesis, the word 'simulation' generally refers to Building Performance Simulation (BPS). In this chapter,

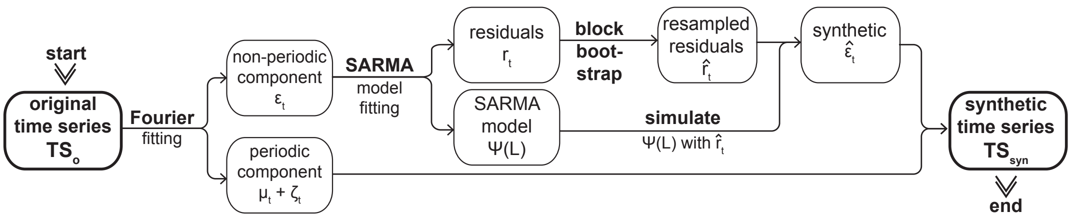


Figure 3.1 – *Generating synthetic weather time series from typical data, without incorporating climate change forecasts.*

preserve the autocorrelation of a time series, i.e., the correlation of a time series with itself; and its cross-correlations with other time series, e.g., the correlations between temperature and solar radiation. The generator produces time series with descriptive statistics that are similar to the historical record, though not necessarily the same. For those synthetic data that incorporate climate change models, verification with recorded data is not so helpful. However, the example typical files used in this thesis provide a validation through historical data. The TMY files used in this thesis tend to be composed of months from the 1970s-1990s (see Section 2.6.1 for details of the generating procedure). When we examine the statistics of the series generated with climate change input, they match those of recorded data from recent decades. That is to say that the variance generating procedure is able to 'update' the older data to match recent decades. This is also the case for the plain synthetic/random files. In this case, they reproduce extremes but do not change the mean. The generation procedure is as follows:

1. Fit a Fourier series to the original time series (y_t). Subtract the fitted terms ($\mu_t + \zeta_t$) and continue with the residuals (ε_t).
2. Fit a Seasonal Auto-Regressive Moving Average (SARMA) model to the Fourier residuals (ε_t). Pick the model based on parsimony (Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Log Likelihood) and residual analysis. Continue with the SARMA residuals (r_t), which are considered to be 'noise'.
3. Shuffle the second-level residuals (r_t) in 3-day blocks, to obtain some desired number of synthetic noise series $\hat{r}_t$ (block bootstrap).
4. Simulate the SARMA model with the shuffled noise to obtain synthetic analogues to the Fourier residuals from step 2 ($\hat{\varepsilon}_t$).

it is also used to indicate the use of a stationary time series model to generate synthetic series.

[2] See the glossary for a description of the Representative Concentration Pathways (RCPs). In this thesis, we will occasionally refer to RCP 8.5 as the "high emissions" scenario, and RCP 4.5 as the "medium emissions" scenario.

5. Add the synthetic residuals ($\hat{\varepsilon}_t$) to the Fourier fits to obtain synthetic temperature or relative humidity values ($\hat{y}_t$)

3.3 Splitting the Time Series into Deterministic and Random Components

A *time series* can broadly be defined as any data collected or observed sequentially over time, such as the Dry Bulb Temperature (TDB) values taken from a typical year file for Geneva shown in Figure 3.2. The procedure used in this thesis breaks up time series into quasi-deterministic (periodic) and random (non-periodic) components. We say *quasi*-deterministic because, while the periodic components are treated as fixed in this procedure, they are obviously not fixed in nature (e.g., the diurnal temperature cycle). Once a periodic series is fitted to the original time series, it remains unchanged until it is added back to the random noise component. A time series with obvious seasonality may be decomposed into periodic and aperiodic components

$$y_t = (\mu_t + \zeta_t) + \varepsilon_t, \tag{3.1}$$

where t is hour of the year, y_t is the original time series, μ_t the low-frequency periodic component, ζ_t the high-frequency periodic component, and ε_t the aperiodic component. The aperiodic or (apparently) random component is the residual from removing the periodic components. In subsequent sections, we will further split the random component into a stationary time series model and a nearly-white noise residual

$$\varepsilon_t = \mu + \psi(L) r_t, \tag{3.2}$$

where μ is the mean of the series, ψ is a polynomial representing a stationary time series model (Section 3.5), and r_t is the residual.

A convenient method for splitting each series into periodic and aperiodic components is frequency domain analysis with the use of Fourier series to represent the periodic or 'deterministic' part. In general, the aperiodic component is expected to be stationary. A handy working definition of a stationary time series is one that has a constant mean, variance, and autocorrelation[3]. Imagine a climate with an annual mean TDB of $\overline{T}_a$. If the time series of TDB in this climate was stationary, then the mean of each

[3] That is, a function that does not vary with time, so-called second-order stationarity.

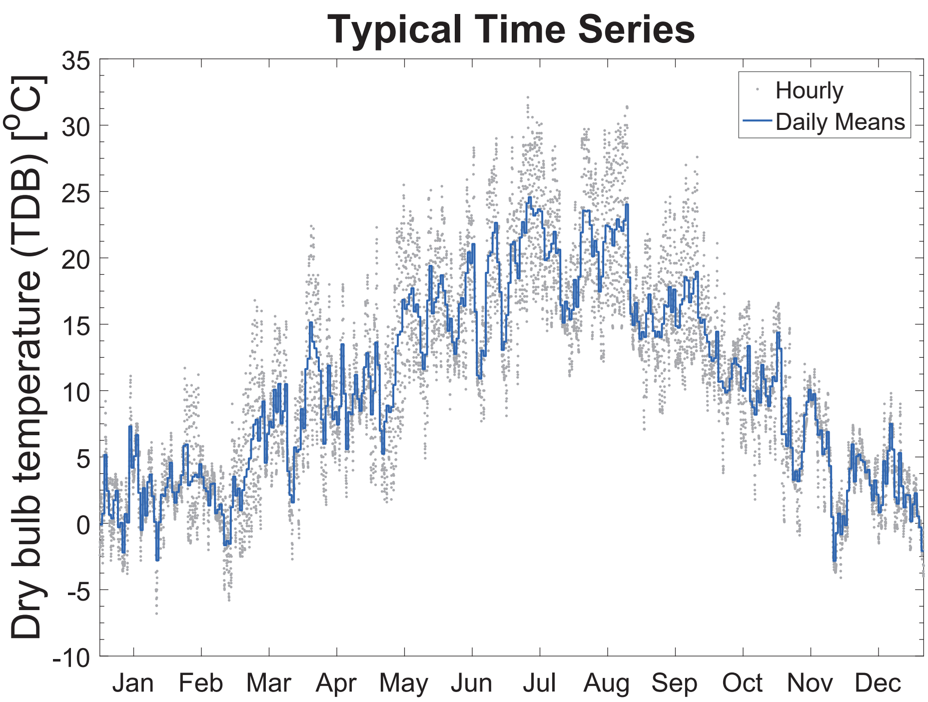


Figure 3.2 – *Dry Bulb Temperature (TDB) time series for Geneva.*

month would be exactly the same as the annual mean and that of every other month, $\overline{T}_{m,1} \approx \overline{T}_{m,2} \approx \cdots \approx \overline{T}_a$. The variances of each month would also be approximately equal. Climate series are often assumed to be stationary over decadal or centennial time spans. For example, in the UKCP09 generator discussed before (Section 2.6.3), stationarity is assumed for each user-defined future 30-year time period (Jones, Harpham et al. 2010). In this thesis, we are only concerned with time series sampled every hour (hourly series), which is the resolution at which building simulation programs typically work. However, sub-hourly time series would also be useful in some cases, like model predictive control or solar power production. This is an interesting line of enquiry for future work.

3.4 Fourier Fitting To Remove Seasonal Trends

We rewrite Equation (3.1) in a Fourier series representation[4]

$$y_t = \alpha_0 + \sum_{i=1}^{m} \left[\alpha_i cos(2\pi\omega_i t) + \beta_i sin(2\pi\omega_i t)\right] + \varepsilon_t, \tag{3.3}$$

where $t = 1, \ldots, n$ is the time vector, $i = 1, \ldots, m$ is the index of the sine-cosine pairs, ε_t is white noise, or normally distributed residuals, ω_i is the (unknown) frequency of the periodic terms, and α, β are unknown coefficients or parameters. A small number of pairs (i.e., sine-cosine pairs at one to three frequencies) constitute a good enough Fourier or spectrum representation of the time series, since we accept that there will be a residual term ε_t. The residual is supposed to be near-white noise, i.e., a sequence with all frequencies, analogous to white light.

Examining the periodogram of the year-long series of hourly TDB values (shown in Figure 3.2), we see the presence of definite peaks at wavelengths 8760 hours and 24 hours, corresponding to 1 and 365 waves a year, respectively. We combine visual observation of the raw data and periodograms with knowledge about the underlying physical processes to choose the least number of periodic components that result in stationary residuals. Thus, the Fourier fit used for Dry Bulb Temperature (TDB) is

$$\begin{aligned} \mu_t + \zeta_t = a_0 &+ a_1 \cdot \cos(N_t) + b_1 \cdot \sin(N_t) \\ &+ a_2 \cdot \cos(N_t/2) + b_2 \cdot \sin(N_t/2) \\ &+ a_3 \cdot \cos(D_t) + b_3 \cdot \sin(D_t); \end{aligned} \tag{3.4}$$

and for Relative Humidity (RH) it is

$$\mu_t = a_0 + a_1 \cdot \cos(N_t) + b_1 \cdot \sin(N_t), \tag{3.5}$$

where t is time, $N_t = \frac{2\pi t}{8760}$, and $D_t = \frac{2\pi t}{24}$. The RH equation has only one term on the left hand side, to denote a low-frequency fit, as opposed to the two terms seen in the equation for TDB.

[4]A brief introduction to Fourier series is in Section A.2.

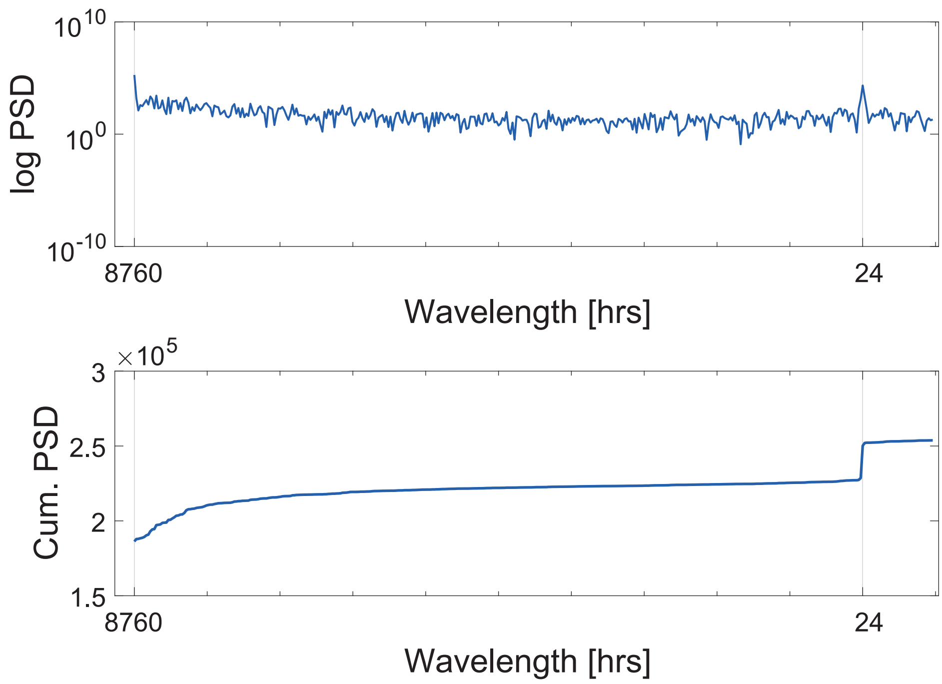


Figure 3.3 – *Periodogram of the Dry Bulb Temperature (TDB) time series for Geneva. Power Spectral Density (PSD) values are on the y-axis and the x-axis is labelled with the wavelengths rather than the frequencies.*

3.4.1 Modifications for Climate Change Forecasts

An innovation in this thesis is the incorporation of climate change forecasts for the creation of *an ensemble of possible future weather years.* Climate change forecasts are available as time series of projected future daily mean values from the Intergovernmental Panel on Climate Change (IPCC) (see Section A.5 for details). The forecasts are distinguished by the future expected concentration of Green House Gases (GHGs) due to policy decisions and technological progress. Thus, for example, RCP8.5 represents business-as-usual while RCP4.5 represents a scenario where emissions peak before 2050 (see Fig. SPM.5a in IPCC (2014b)). The temperature trend depends on the Representative Concentration Pathway (RCP) selected, i.e., the future concentration of GHGs one wishes to test. In addition, there are several combinations of Global Climate Models (GCMs) and Regional Climate Models (RCMs) to choose from.

The future time series of daily means are used as replacements for the low-frequency

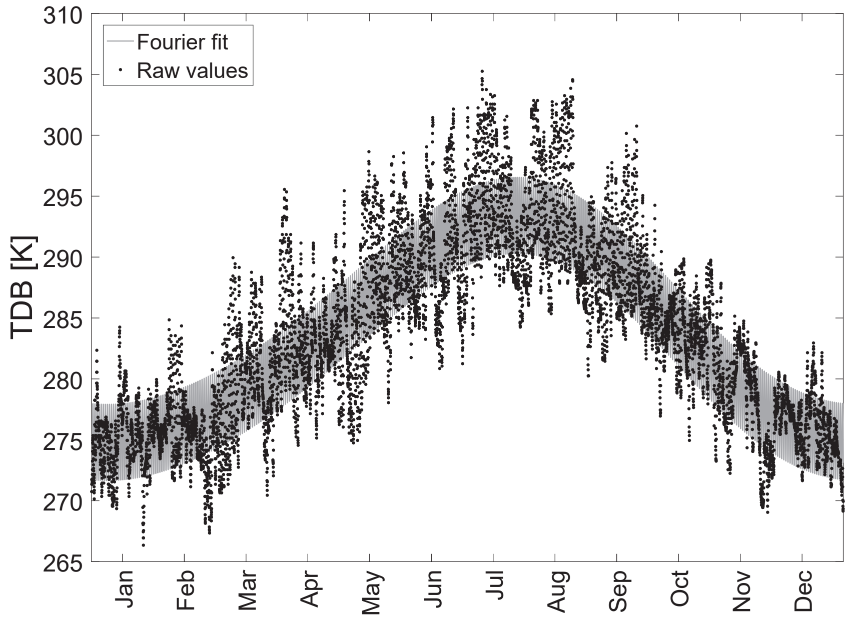


Figure 3.4 – *Raw TDB values (black dots) plotted against the Fourier series (grey lines) for Geneva.*

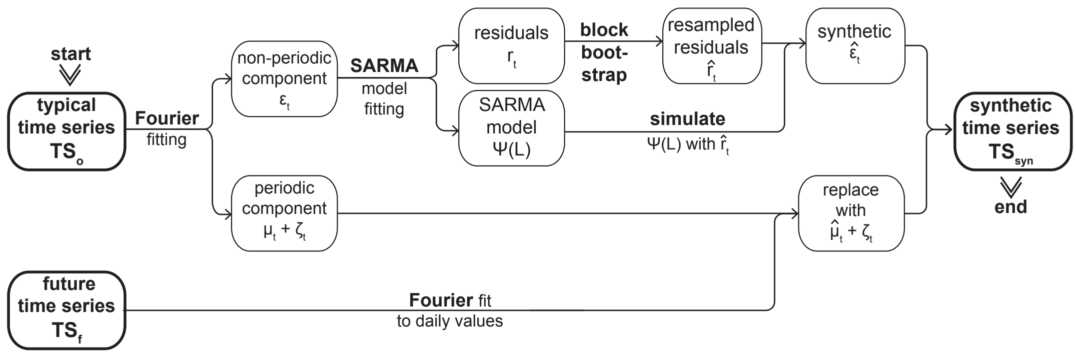


Figure 3.5 – *Using future forecasts of daily mean values with random noise series. The low-frequency periodic component of the original series is replaced with the Fourier fit to future daily values obtained from climate change models.*

Fourier term μ_t in the recombination step described in Section 3.7 and fig. 3.5. Recall, from Equation (3.4), that the 'deterministic' fit contains two or three trigonometric pairs, but we are only replacing the annual-wavelength pair of sine-cosine terms. We assume that the daily fluctuation components will not change in the future. By

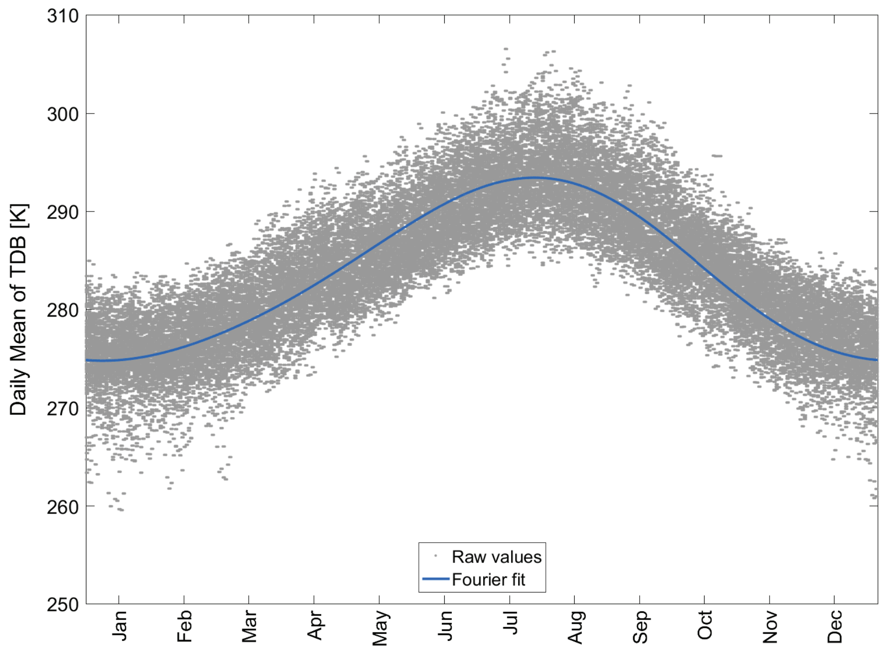


Figure 3.6 – *Raw future daily mean TDB values and the annual Fourier fit. Data from Geneva.*

this we are implicitly assuming that the amount of cloudiness will not change in the future, since that is the primary driver of the diurnal temperature swing. The future daily mean values are shown in Figure 3.6, along with the low-frequency Fourier fit. Examining the modelled future daily mean Humidity Ratio (W) values downloaded from the CORDEX website, we found that its characteristics are not different from recorded or typical data. That is to say that the predictions for Geneva do not forecast an appreciable change in the wet bulb characteristics of the climate. Therefore, while the climate change forecasts are applied both to the TDB and RH time series, the effect is only noticeable in the TDB series.

3.4.2 Characteristics of the Residuals

This relatively parsimonious Fourier fitting successfully removes the annual and daily harmonics, where they exist in the original meteorological data, leaving stationary time series with zero means. Figure 3.4 shows the raw values and the periodic series. Figure 3.7 shows descriptive plots and Figure 3.8 a periodogram of the residual from fitting Equation (3.4) to the TDB data. These plots show that the daily and annual harmonics have been stripped out completely by subtracting the Fourier terms. Slightly raised Power Spectral Density (PSD) values around (but not at) $\omega = 0.0417$, or approx. 24 hours, correspond to leakage around the daily fluctuation. High PSD values at 12 hours correspond to unknown waves of sub-daily length, and are probably due to leakage as well. The rest of the periodogram is relatively flat, which means it is closer to the white noise spectrum than the previous periodogram (Figure 3.3). The sub-plots of Figure 3.7 also tell us that the residuals sufficiently resemble a stationary "white noise" series[5]. The plot of raw values over time (top-left) shows a roughly constant mean and variance, while the qqplot suggests that the data belongs to a normal distribution. In the correlograms, the horizontal lines around zero represent the cut-offs for significance. If the (P)ACF value at a given lag falls between these lines, it is insignificant. The correlograms (bottom) show that, while the series might be weakly stationary, there is structure that can be exploited. To do that, we need stochastic models for stationary time series.

The picture is slightly different for the RH series residuals presented in Figures 3.9 and 3.10. The periodogram shows a relatively high amount of power at and around 24 hours, and around 720 hours (about 1 month). The apparently strong signal at 24 hours, i.e., a sine-cosine pair with a period of one day, is misleading; adding a second Fourier term in this step did not improve the residuals. We know of no physical reason why humidity should vary as a diurnal sine wave, except that the quantity RH is somewhat correlated to temperature (Table 3.3)[6]. The descriptive plots are similar to those for TDB, i.e., they point towards a sufficiently stationary residual with auto-correlation structure.

[5] If the residuals from a fit are close to white noise, then the fit is good.
[6] Recall that the *relative* humidity is a function both of the moisture content of the air and its temperature.

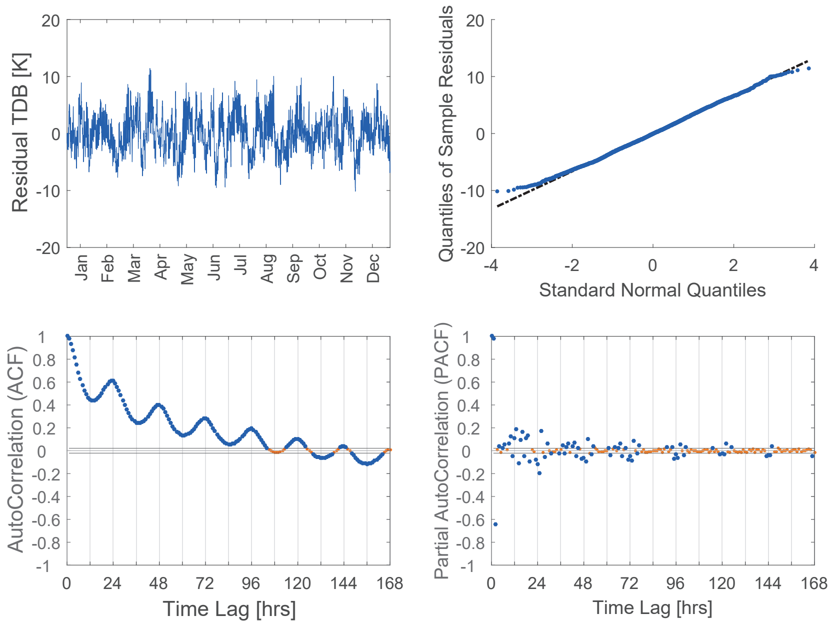


Figure 3.7 – *TDB, Geneva. Clockwise from top left: raw residuals; qqplot of residuals; correlograms of residuals.*

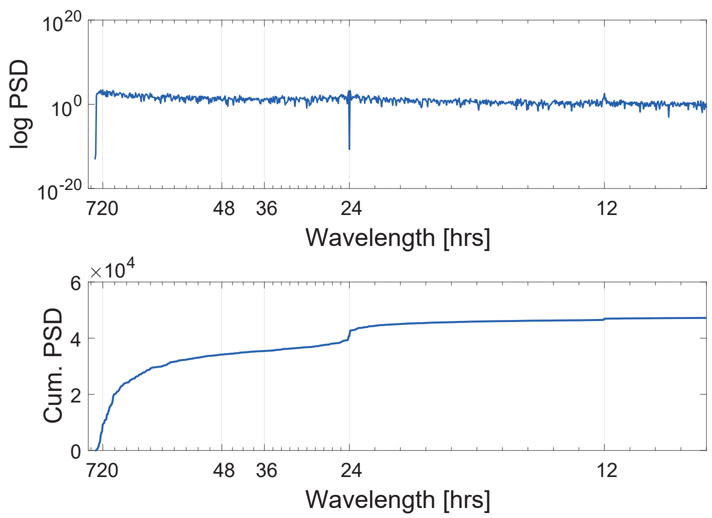


Figure 3.8 – *TDB, Geneva. Raw and cumulative Power Spectral Density (PSD) of the residuals from the Fourier fit.*

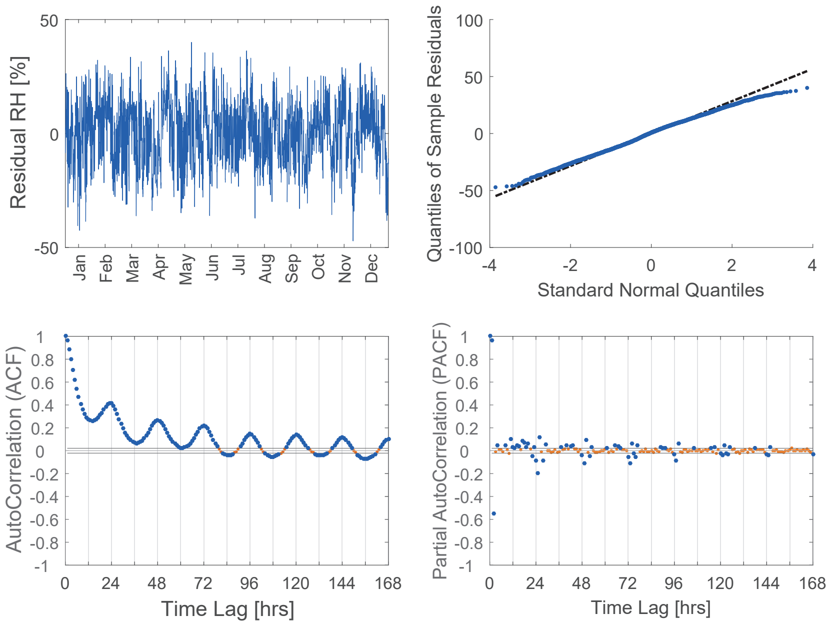


Figure 3.9 – *RH, Geneva. Clockwise from top left: raw residuals; qqplot of residuals; correlograms of residuals.*

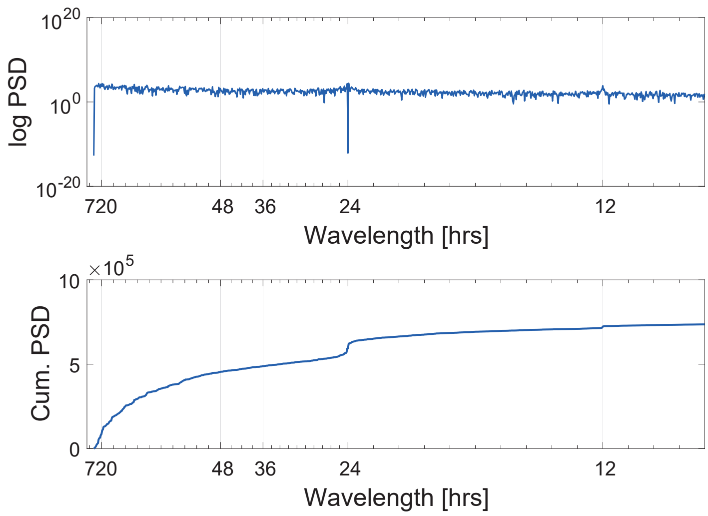


Figure 3.10 – *RH, Geneva. Raw and cumulative PSD of the residuals from the Fourier fit.*

3.5 Stationary Time Series Models for Random Components

Details of various model types can be found in Appendix A, time series analysis textbooks like Box, Jenkins et al. (2008), Christensen (1991) and Cryer and Chan (2008), and the documentation of software like the ones we used: the `arima`, `estimate`, and `infer` functions in MATLAB®(The MathWorks, Inc. 2015); and the `forecast` package in R (Rob J. Hyndman and R Core Team 2015). We now explain the use of these models for stationary time series. The model selection process used in this work is based on the Box-Jenkins methodology (Box, Jenkins et al. 2008). A brief introduction to Auto-Regressive Moving Average (ARMA) and other time series models is in Section A.3.

One of the chief concerns in choosing various models for this procedure has been their interpretation. We have tried to choose only those methods/models that can be interpreted physically. For example, an auto-correlation term for temperature at time lag 24 hours is intuitive, but one at lag 23 is not. In a purely data-driven procedure, lag 23 is as good a candidate as lags 24 or 25. If we use automatic fitting procedures, the 'optimal' Auto-Regressive (AR) terms are sometimes at lags 23 or 25. The use of seasonal and non-seasonal terms in combination ensures that, for example, a model with 24-hour seasonal lags and 1- to 3-hour non-seasonal lags accounts for the terms at lags 24, 25, 26, and 27 hours. We generally use simple models that reflect our biases in interpreting them[7]. Also for ease of interpretability, we eschewed the use of differencing for stationarity. In any case, it does not confer much advantage in our application. The series tend to be close enough to stationarity just by removing a Fourier fit (Figures 3.7 to 3.10). The use of differencing complicates the interpretation of the residual and its simulation. Hence, the differencing factor in the time series models we use should always be zero[8].

To choose the best model for a climate, we used information criteria (AIC and BIC) and log-likelihood values. In this case, we knew something about the physics of the underlying time series and were able to exploit that knowledge to pick the relevant *class* or *type* of models. It is only the estimated coefficients, or model parameters, that change between climates. Since the coefficients are empirical, i.e., they describe the influence of model lags *based on the data from that specific climate*, they are easily calculated from a typical file. The decision to include/exclude certain lags, for example, may be made by an expert to suit their case. The stationary time series models and Fourier fits used in this thesis are *time-invariant*. That is to say that the coefficients

[7] The length of the actual solar day varies slightly, with a quarter day going 'missing' every year. Precise timekeeping is not a worry in building simulation, so we will leave considerations of leap years and seconds aside and use a 24 hour day. The influence of a leap day on the annual weather pattern is not detectable.

[8] An ARMA model with differencing is called an Auto-Regressive Integrated Moving Average (ARIMA) model.

and terms are fixed for a time series, or that the coefficients and terms themselves are not functions of time. The model coefficients are calculated with the entire year-long time series used in this procedure (8760 hours).

3.5.1 Correlation Estimation

The correlograms shown in Figures 3.7 and 3.9 give empirical estimates of the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) of the time series. “Empirical” because they are calculated from the data at hand, rather than from knowledge about the underlying function. The values are plotted against time lags of 0-168 hours. The ACF is a measure of the influence of the values at a particular lag on values at the current time step. The PACF is the auto-correlation left over after adjusting for any effects of underlying linear structure. Software usually calculates the PACF by successively fitting AR models of order $p = 1, 2, \ldots$. If the ACF cuts-off abruptly, while the PACF decays gradually, then we should expect a Moving Average (MA) model. Conversely, if the ACF decays gradually and the PACF cuts off abruptly, an AR model is probably appropriate. If both decay gradually, then an ARMA model may be called for. A SARMA model is an extension of this concept to account for significance at periodic lags. For example, the ACFs in Figures 3.7 and 3.9 [bottom left] show significant correlations at multiples of 24.

3.5.2 Selecting a Model

A stationary time series model is selected by estimating the significant lags and corresponding coefficients. In our implementation, the lags tested in the models are within very narrow ranges (0-4), and the actual coefficient for each lag is calculated for the climate at hand using maximum likelihood estimation, details about which may be found in textbooks like Christensen (1991) and Davison (2003).

The overarching principle of model selection is parsimony. That is, given two models that explain the same amount of the underlying structure, pick the one that has less parameters. Log likelihood values, characteristics of the residuals, expert knowledge, robustness of the model, and generalisability serve to pick a model in a reasonably automated manner. To reduce the amount of expert intervention we used three quantitative measures, the log likelihood and two information criteria

$$\text{AIC} = -2 \cdot \ell(\theta) + 2 \cdot n_p, \text{BIC} = -2 \cdot \ell(\theta) + n_p \cdot \log(n_o), \tag{3.6}$$

where $\ell(\theta)$ is the log likelihood, n_p is the number of parameters in the model, and n_o is the number of observations. The information criteria should be as small as possible, because they penalise extra parameters while rewarding likelihood. However, the automated selection of models based on these quantities is not foolproof. When comparing several models, the largest maximum likelihood values will probably be from models that are over-fitted, i.e., have unnecessary terms (based on a visual inspection of the correlograms). For example, the correlograms in Figure 3.11 (for New York LaGuardia airport) indicate an AR model, with perhaps an additional seasonal component. However, the log likelihood values were consistently higher for models with additional MA components. AIC and BIC are better for selecting models because they penalise overfitting.

3.5.3 The Selected SARMA Model

A (multiplicative) SARMA model is a parsimonious extension of ARMA models to include the concept of *seasonal* dependence. Note that the idea of seasonality should not be confused with the *seasons* (like spring or summer), which are a cultural construct. A SARMA model consists of a combination of seasonal and non-seasonal AR and MA terms. The seasonal terms include every Q^{th} or P^{th} past value, e.g., every 24 hours back from the present, and the non-seasonal terms refer to the last q or p terms from every periodic lag, e.g., 1-4 hours before the present and every 24 hour step back. In this application, the seasonal terms ensure that values from the same hour on previous days have an influence on the current value, in addition to values from immediately preceding hours.

The SARMA models we use are

$$\text{SARMA } (p,q) \times (P,Q)_s, \quad \text{or,} \quad \phi(L)\Phi(L)\varepsilon_t = c + \theta(L)\Theta(L) r_t, \tag{3.7}$$

where s is the seasonality or periodicity; the upper-case versions of $\phi(L)$ and $\theta(L)$ represent the seasonal terms, ε_t is the residual after removing the Fourier terms (Equations (3.4) and (3.5)), and r_t is the residual after fitting a SARMA model to that residual series. For details about the notation and background on the model

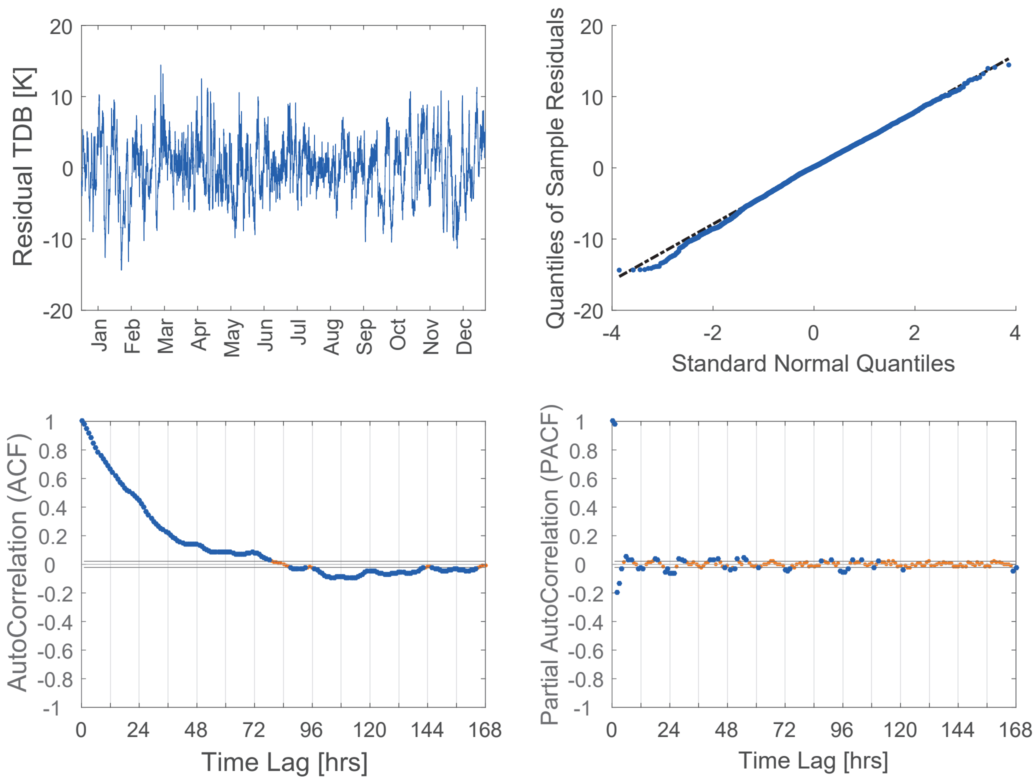


Figure 3.11 – *Residuals from de-seasonalised TDB from New York LAG. The correlograms indicate an AR model, with weak to no Seasonal Moving Average (SMA) coefficients.*

components, see Section A.3. The seasonal and non-seasonal differencing factors are both zero. For example, the model for the data shown in Figure 3.2 is

$$\text{SARMA}\,(4,2)\;\times\;(1,1)_{24},$$

where the coefficients to be estimated are $\{p_1, p_2, p_3, p_4, q_1, q_2\}$ and $\{P_{24}, Q_{24}\}$. In our work, we will use the Box-Jenkins notation, where the P and Q of Equation (3.7) are specified as multiples of the seasonality s[9].

[9] This is how the `forecast` package in R also specifies models. The `arima` and related packages in MATLAB, however, use a notation where the seasonal lags are specified as absolute numbers, e.g., 12, 24, rather than as multiples of the seasonality.

3.5.4 Characteristics of the Residuals

After selecting a model for a given climate, e.g. SARMA $(4,2) \times (1,1)_{24}$ from Geneva TDB, we examine the residuals (Figure 3.12). Using the same plots as Figure 3.7, we examine the data for *stationarity*, *homoscedasticity*, *Normality*, and a lack of discernible correlations. The top-left sub-plots in Figures 3.12 and 3.13 show that the residuals are probably stationary and homoscedastic. The qqplots at the top-right do not show a favourable comparison with a Normal Probability Density Function (PDF). The correlograms on the bottom show that there is a trace amount of structure left over, at $k \leq 72$ hours. This suggests the use of 72-hour blocks during the resampling procedure ($R_1, R_2, \ldots, R_N$ in Section 3.6.1). Significant lags were usually seen below 72 hours in the climates tested, while the probability distribution tended to depart from Normality sometimes, as here.

Cryer and Chan (2008) argue that in an ARMA model, there is always some dependence of the variance of the residual on the values of the residual itself, i.e., *hetero*-scedasticity. In addition to the SARMA model, we tested Auto-Regressive Conditional Heteroscedasticity (ARCH) and Generalised Auto-Regressive Conditional Heteroscedasticity (GARCH) models for the variance – combining conditional mean models with conditional variance models. However, the fit did not improve for either of the time series under consideration, TDB or RH, so we abandoned this line of inquiry.

The final step in the process of generating some desired number of 'synthetic variants' is to simulate this SARMA model. That is, generate new samples of the $r_1, r_2, \ldots, r_n$ series to plug back in to Equation (3.2). Normally, the noise or innovation series used to simulate the conditional mean models are Gaussian white noise or a Student's t-distribution[10]. This thesis uses a customised distribution of innovations – bootstrapped residuals[11]. That means that our statistical analysis is non-parametric, whereas sampling from a distribution would imply that we know the distribution of the underlying random variables $R_1, R_2, \ldots, R_n$ (a parametric approach).

[10] The weather generators described in Section 2.6.3 all add Gaussian noise.
[11] See Davison and Hinkley (1997) for details.

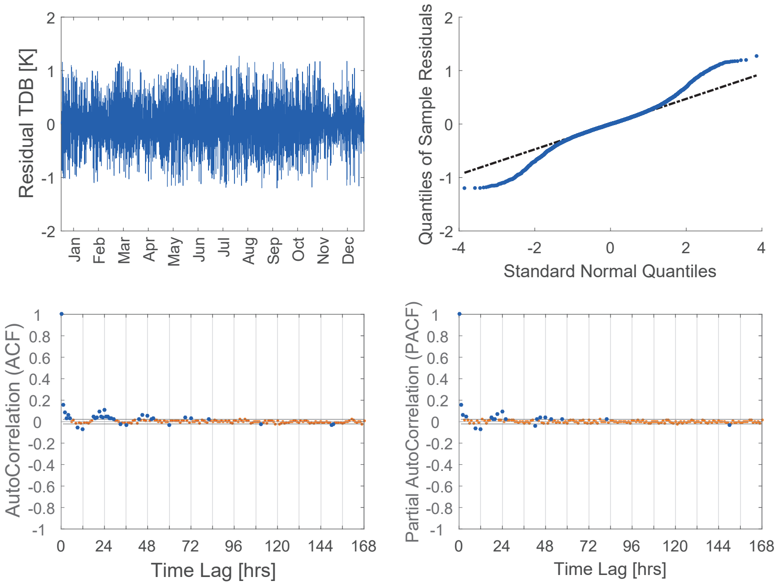


Figure 3.12 – *Residuals (r_t) from fitting a SARMA model to the 'de-seasonalised' hourly TDB time series for Geneva (ε_t).*

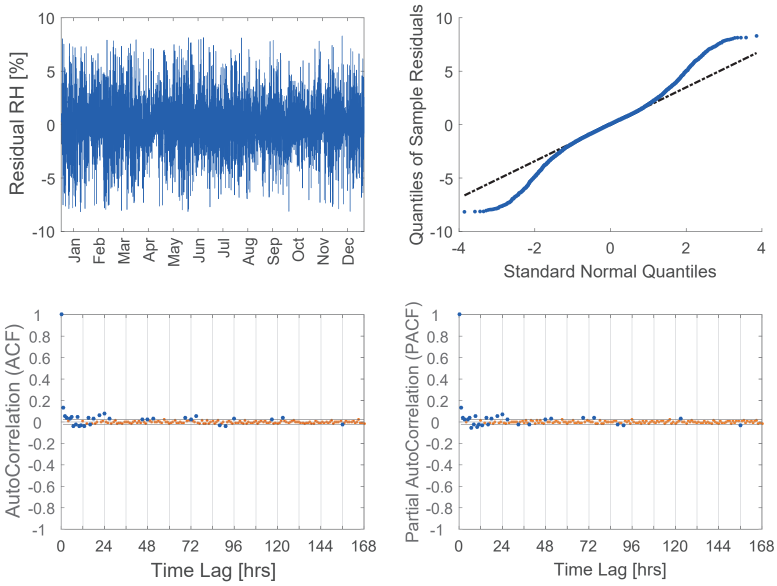


Figure 3.13 – *Residuals (r_t) from fitting a SARMA model to the 'de-seasonalised' hourly RH time series for Geneva (ε_t).*

3.6 Introducing Noise for Variation

3.6.1 Resampling and Subsampling

Resampling and subsampling are methods to improve estimates of statistical measures like the mean and confidence intervals for small samples. Well-known resampling and sub-sampling methods include the bootstrap and the jackknife. Essentially, "the bootstrap method amounts to treating your observed sample as if it exactly represented the whole population" (Politis 1998). These methods do not rely on knowing particular parameters describing the shape or distribution of the population or sample (e.g., mean, standard deviation, etc.). They also avoid the restrictions imposed on a parametric random model by a small sample size (the TMY in our case) and long sample runs. Politis (ibidem) contends that resampling methods are more useful in non-parametric situations where a model is not available, and the data has to "do all the talking". See Davison and Hinkley (1997) and Politis (1998) for a discussion of how bootstrap, jackknife, and similar methods can be used for valid Monte Carlo estimation of population parameters like the mean and variance.

Resampling/subsampling methods cannot be used directly on time series that show a high degree of seasonality or correlation, since they do not account for underlying structure. For example, temperature on a July night is highly correlated to a July day, but not necessarily to a January night. A sampling run cannot distinguish between day and night temperatures or summer and winter – a fatal problem for weather data. We avoided this problem by using the series of model fits, described before, to ensure that the series to be bootstrapped contains as little underlying structure as possible. See Politis, Romano et al. (1999) for a discussion of other options for subsampling from time series.

Bootstrapping and jackknifing are often used to get an estimate of the bias (systematic error) and variance (random error) of some estimator T (e.g., sample mean) of a population parameter θ (e.g., population mean). In our application, we are not interested in the parameters as such, just in creating variance (in the innovations for a SARMA filter) without changing the underlying (unknown) distribution of the innovations. In a non-parametric bootstrap, the simulated samples $R_1^*, R_2^*, \ldots, R_N^*$ may be taken with or without replacement from the original time series. In the flavour of bootstrap/resampling that we use, the *residuals are resampled per month and without replacement*. This amounts to a shuffling of the data, effecting as little a change in the underlying structure remaining in the residuals as possible. This is an unusual implementation of the jackknife, or subsampling, as opposed to the bootstrap, or

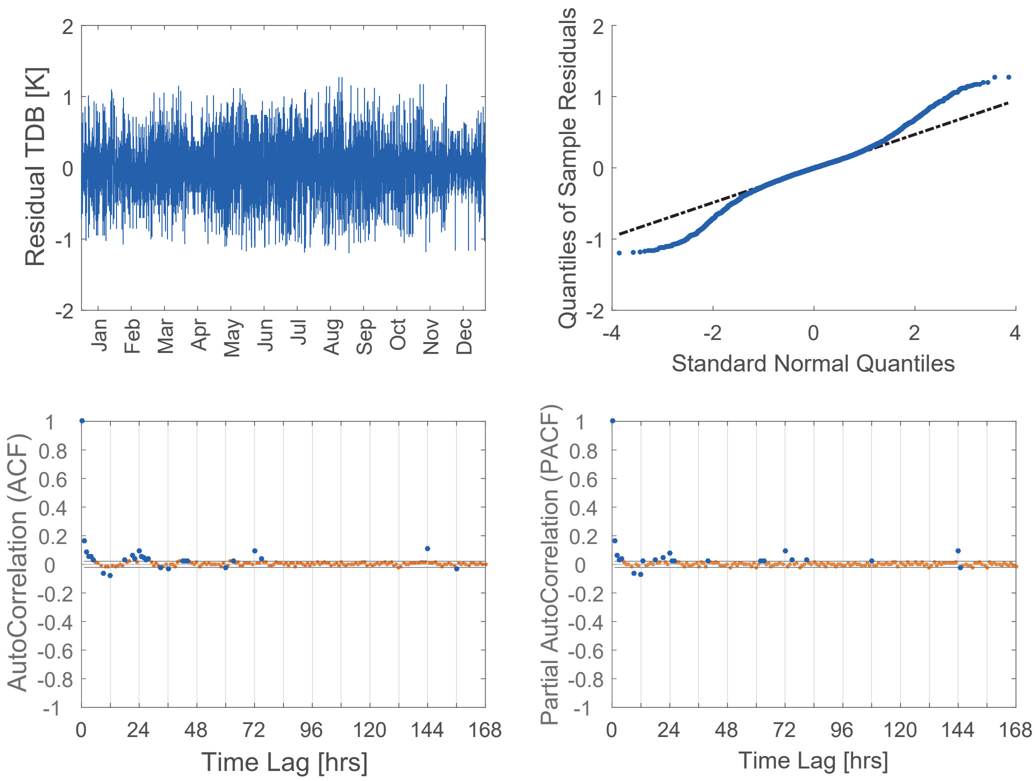


Figure 3.14 – *A randomly selected series of $\hat{r}_t$ values from the bootstrapping step – TDB, Geneva.*

resampling. Tukey's original *jack-knife* involves the use of sub-samples of size b, $b < N$ (Politis 1998), whereas the size of sub-sampled series in this application is $b = N$. Here b is the size of the sub-sample, while N is the size of the sample.

We decided against bootstrapping individual r_t values because the structure at lags $k \leq 72$ hours is still present in the candidate series (Figures 3.12 and 3.13). It is necessary to maintain this ACF not only to remain faithful to the original series but also to return a physically valid time series of temperatures at the end of the process. Using a 'block bootstrap' instead of shuffling individual points has a better chance of preserving short-term auto-correlations (Davison and Hinkley 1997; Politis, Romano et al. 1999; Politis 1998). The bootstrap is done using 3-day blocks, which is a compromise between introducing enough variance without introducing bias. It is also the block length suggested by Magnano, Boland et al. (2008). We chose a 3-day block after exploring different options (1-7 days) to see which block length reproduces

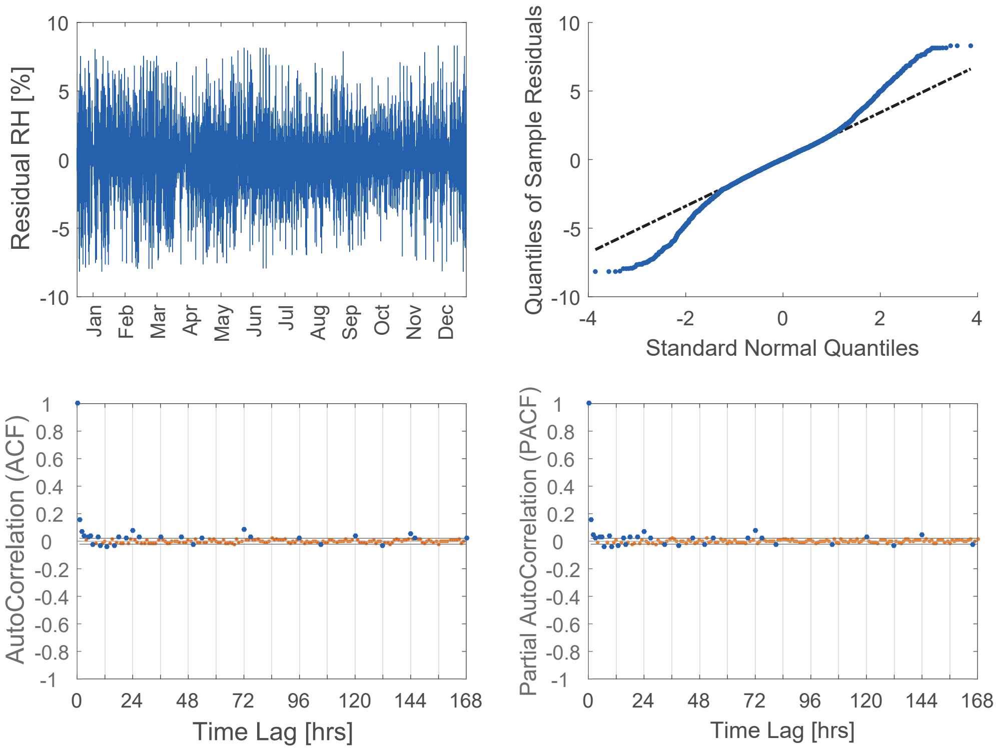


Figure 3.15 – *A randomly selected series of $\hat{r}_t$ values from the bootstrapping step – RH, Geneva.*

the auto-correlograms and partial auto-correlograms best. The results from different block lengths are not substantially different upon visual inspection. Unsurprisingly, any block length preserves the intra-block ACF. The behaviour up to 7 days is not consistent, but the differences between the correlograms of the original residuals and shuffled series are very small. We went with the smallest valid block size, based on preserving the ACF and PACF plots, to get as much variability among the synthetic series as possible. This choice is the classical one between variance and bias – the former usually cannot be increased without also increasing the latter. Plots of one synthetic noise series for TDB and RH each are presented in Figures 3.14 and 3.15, respectively. Individual series will usually reproduce the statistical characteristics of the original series slightly worse than the *ensemble*, but calculating/plotting the ACF and PACF of an 8760×100 series was too much for the computer we used.

Another restriction imposed on the bootstrap was to subdivide the blocks by month and limit the shuffling to blocks within a month. For example, new series of noise blocks for January were sampled only from January noise and not, say, July. This helped to preserve any residual sub-yearly correlation. Despite all of these restrictions, the bootstrapped/simulated residuals produce enough variation. Let the set of residual blocks

$$\mathscr{B} = B_1, B_2, \ldots, B_N, \text{where} \begin{cases} n = 10 & \text{for months with 30 days or less} \\ n = 11 & \text{for months with 31 days,} \end{cases}$$

and each B_i contains 72 hours of residuals. Then, each new series of simulated residuals is $\mathscr{B}_j^* = B_{j,i^*}$, where j is the index of a residual set, and i^* is the index of element blocks in that residual set. Since the index i^* is any permutation of the original index $i = 1, 2, ..., n$, the number of variations possible is

$$N! = \begin{cases} 3628800 & \text{for } N = 10 \\ 39916800 & \text{for } N = 11. \end{cases}$$

3.6.2 Simulating the SARMA model

We simulated the SARMA model using the `forecast` function in R (Rob J. Hyndman and R Core Team 2015). A sample R script is given below, generated from a function we wrote in MATLAB, `SimArimaR.m`[12]. The script takes the resampled residuals and simulates the SARMA model that has been passed to it with these resampled values serving as the noise input. Usually, the noise input are samples from a Normal distribution. There are sometimes differences in the model coefficients calculated by the fitting functions in MATLAB and in R, but these are minor. This is due to the different optimisation algorithms used in the Maximum Likelihood Estimation (MLE) step.

[12] Available with the archive copy of this thesis on **infoscience.epfl.ch**. Comments have been added for presentation, and the script reformatted slightly from the one actually used in the calculations.

```
# This R script is written automatically by a function written
# in MATLAB (see infoscience.epfl.ch).

# Load forecast
library(forecast)

# Read in the custom innovations.
CustomInn = read.csv(file.path("Inn.csv"))
# And the raw values.
IncomingX = read.csv(file.path("Raw.csv"))

# These values have been written in by MATLAB.
periodicity = 24; ARp = 3; MAq = 4; SARp = 1; SMAq = 1

# Try fitting an ARIMA model. Sometimes the MLE does not
# converge with the CSS-ML option.
try( PsiModel <- Arima(IncomingX, order=c(ARp,0,MAq),
    seasonal=list(order=c(SARp,0,SMAq), period=periodicity),
    method = 'CSS-ML', n.cond=0) )

# If the MLE did not converge, then try the CSS option.
if (!(exists('PsiModel'))) {
    try( PsiModel <- Arima(IncomingX,order=c(ARp,0,MAq),
        seasonal=list(order=c(SARp,0,SMAq), period=periodicity),
        method = 'CSS', n.cond=0) ) }

# Each series is 8760 values long, and there are a 100 series.
N = 8760; npaths = 100;

# Simulate each series.
tsOut <- array(0,c(N,npaths))
for (s in 1:ncol(tsOut)) {
    tsOut[,s] <- simulate.Arima(PsiModel,nsim=N,innov=CustomInn[,s])
}

# Write out the data in a data frame.
DataOut = data.frame(tsOut)
write.table(DataOut, file="Sim.csv", col.names = F,
row.names = F, sep = ",")
```

3.7 Recombining Random and Deterministic Components

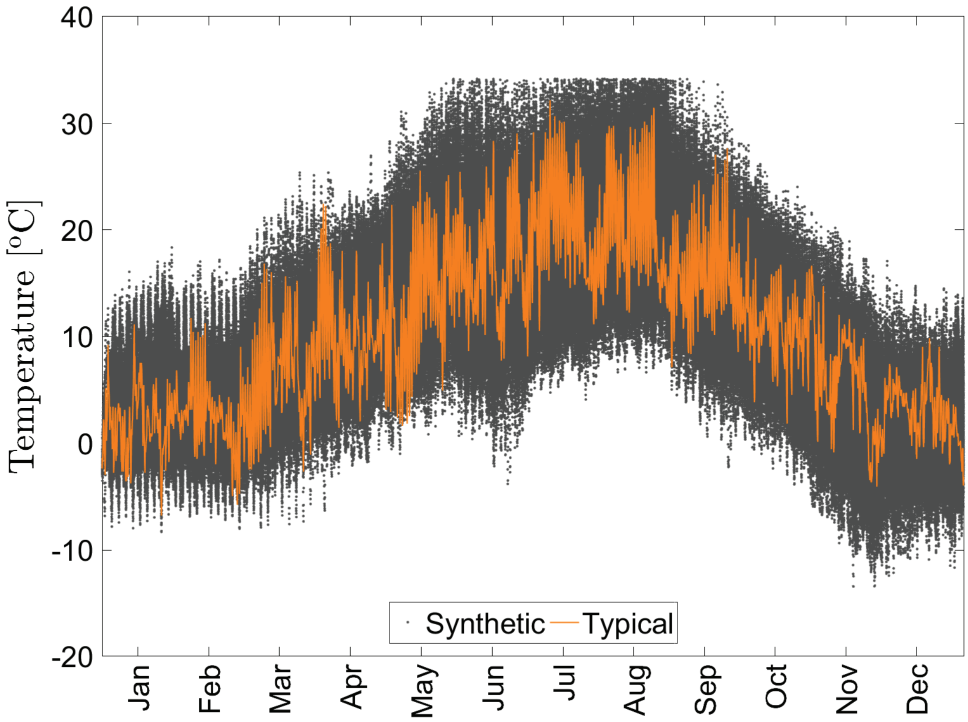


Figure 3.16 – *Synthetic values for TDB, Geneva, plotted with typical values in the foreground. The time series seem to cut-off at about ~35°C. This is a result of the post-processing described in Section 3.8. The maximum temperature in any year can occur between May and September, but does not exceed ~35°C.*

The original decomposition, Equation (3.1), is restated as

$$y_t = (\hat{\mu}_t + \zeta_t) + \hat{\varepsilon}_t, \tag{3.8}$$

where the quantities with hats are synthetic replacements, so $\hat{\mu}_t$ is the low-frequency Fourier fit, and $\hat{\varepsilon}_t$ is the simulated SARMA noise. Plotting the synthetic temperature series against the typical values forms an estimate of variation of temperature at every hour (Figure 3.16). Note that the individual series are all plotted together to achieve this effect. These are not intervals, and intervals cannot be used in current building simulation programs, which only work with a single value per hour. By and large, the expected ranges seem to be slightly larger on 'top'. That is, the mass of the temperature values is shifted slightly towards higher temperatures. The fortuitous outcome of this asymmetry, however, is that the synthetic weather based on an old file (usually 1970s-1990s) mimics temperature trends from the last two decades very well.

3.8 Post-Processing the Synthetic Series

The generation procedure almost always produces unacceptable values (e.g., 100°C), which may be removed with expert input. The difficulty here lies in assessing what an 'acceptable' extreme is: if 40°C has not been recorded in Geneva in the past 60-odd years, does that mean it will never occur in the future? Not necessarily. As such, we do not take a position on the matter and let the user decide on what they consider to be possible, but highly improbable, or ridiculous. We recommend cleaning not just outlandish values (e.g., 50°C in Geneva), but also unrealistic changes in values (first difference of time series). Once again, whether a change is realistic or not is subjective and context-dependent.

To remove unrealistically large values of the first derivative, we use smoothing (local polynomial fits like moving averages). As the name implies, the result would be a 'smoother' signal. Magnano, Boland et al. (2008) smooth the edges of the bootstrapped blocks to ensure that the edges of the blocks do not have an unacceptably large difference between consecutive values. We used the maximum/minimum first difference seen in the recorded data as the limits of what is an 'acceptable' hourly change of temperature. First differences outside these bounds set by the typical file are considered outliers. The actual temperatures corresponding to these outliers were then replaced with interpolated values. We chose not to use a 'global' smoothing, since that removed a significant amount of variance. While some users may wish to implement a global smoothing filter for a generated time series, we do not recommend it.

We also censored the synthetic time series using z-scores, described by Equation (4.2). This procedure is loosely based on one recommended in *NIST/SEMATECH e-Handbook of Statistical Methods*. There are two censoring steps: one for simulated residuals (Section 3.6.1), and one for final synthetic values after recombination (Section 3.7). Censoring was necessary since the innovations and actual values can at times exceed 100°C. By itself, the z-score does not indicate that a particular data point is an outlier. Rather, an arbitrary cut-off point must be decided. The advantage of using z-scores is that instead of imposing arbitrary limits on the raw values of a parameter, which are highly climate or context dependent, it is possible to use standardised values and cut-offs. This helps to maintain consistency across climates and parameters. In our case, we found that choosing the larger of the 99.9 and 0.1 percentile values is sufficiently conservative to remove outrageous values (like 100°C or -100°C) but not so conservative as to remove extremes. The point of using both ends of the range of a time series is to prevent the high extremes causing the censoring of low extremes, or vice versa. When we use the maximum of the 99.9 and 0.1 percentiles (compared using absolute values), we allow for a larger range. A notable exception to this were the simulated TDB values for New York, for which we used the 99th and 1st percentiles. Censoring too, is an arbitrary choice, and the generation of weather files is only moderately affected if this cleaning is not carried out. We looked at various cut-off values, and could not arrive at a conclusively universal one, because we do not take a position on *which extreme is too extreme*. We expect that visual inspection or expert opinion is as good as hard-coded checks in the generator. Most building simulation programs have their own cut-offs for valid values[13].

3.9 Special Treatment for Solar Quantities

... and you run to catch up with the sun but it's sinking
Racing around to come up behind you again
The sun is the same in a relative way, but you're older
Shorter of breath and one day closer to death.

Pink Floyd, *Time*
(The Dark Side of the Moon)

Initially, we attempted to fit Fourier series and stationary models to the GHI series, and a derivative series called the clearness index, K_t[14]. However, we found that the solar time series are both difficult to work with and of dubious quality in many TMY files. As a result, we adopt an approach that exploits the close relationship between the average daily temperature and daily sum of GHI (Figure 3.17), similar to a nearest-neighbour bootstrap. Nearest-neighbour bootstrap has previously been explored for meteorological series, for example by Lall and Sharma (1996) to predict hydrologic data.

Once the synthetic TDB series have been created, the data is separated by month. Then k_{nn} days are selected from a month in the TMY file based on the 'closeness' of their mean TDB values to the mean TDB of each synthetic day in that month (i.e., for

[13] Since we use daily mean TDB in the generation of synthetic future GHI, censored values are easier to work with.
[14] An approach adopted by many of the weather generators presented in Section 2.6.3.

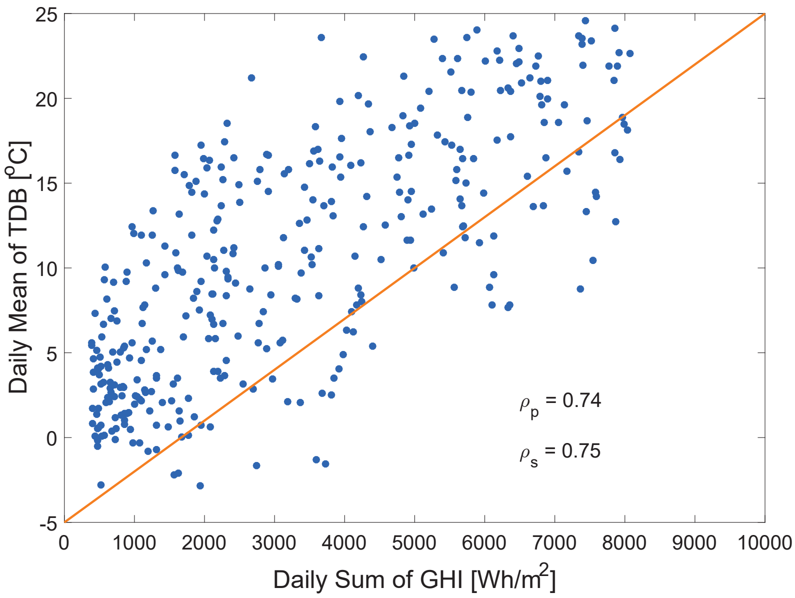


Figure 3.17 – *Daily means of TDB and daily sums of GHI for Geneva. The correlation coefficients are quite high, with Pearson's $\rho_p = 0.74$ and Spearman's $\rho_s = 0.75$.*

each synthetic year). The distance is calculated based on the Mahalanobis distance because it is meant to be unaffected by correlation, though this choice is not fixed and another distance metric could be used. In our implementation, $k_{nn} = 10$, so 10 nearest neighbours from a TMY month are selected for each synthetic day in a month. In the next step, 9 out of 10 neighbours are discarded randomly, and the index of the neighbour selected is stored. Once a TMY day is assigned to every synthetic day, we have pairs of days whose daily mean TDB are similar and which come from the same month. The hourly GHI profile of the TMY day in a pair becomes the hourly GHI profile of the corresponding synthetic day. The corresponding hourly DHI and DNI profiles are selected with the same indices.

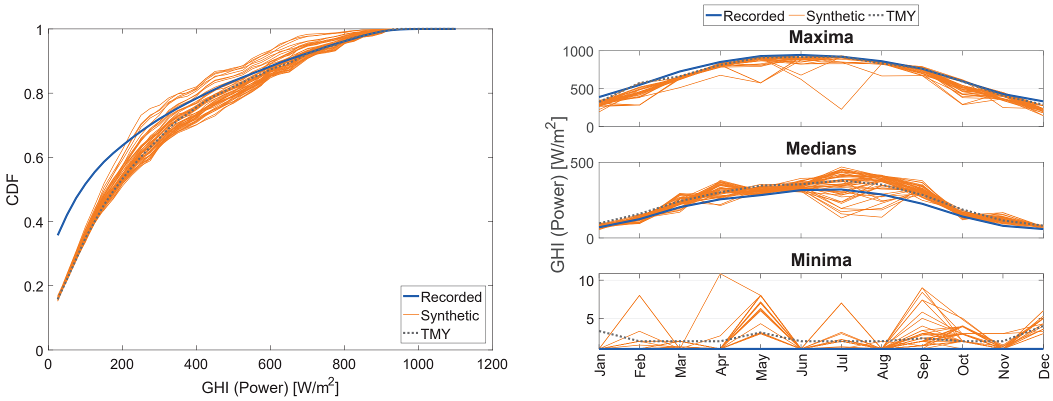


Figure 3.18 – *Synthetic GHI time series for Geneva. Neither the distribution nor the extents of the synthetic series are very different from the TMY or recorded data.*

3.10 Examining the Synthetic Weather Series

For I have known them all already, known them all –
Have known the evenings, mornings, afternoons,
I have measured out my life with coffee spoons;

T.S. Eliot,
The Love Song of J. Alfred Prufrock (1917).

Recall that the purpose of creating the time series is to obtain probable future weather time series, simulation with which would give estimates of the likely ranges of outputs, e.g., energy usage or indoor temperature. We found that the generated data sets are broadly representative of recorded data, despite the fact that long term records are not used in the generation process. We now proceed to assess the results of the synthetic weather data generation using criteria from Boland (1995), Hansen and Driscoll (1977), Lund (1995) and Magnano, Boland et al. (2008). The comparisons were all made using recorded/measured and TMY data. The recorded/measured data used in all the comparisons is from the period 1955-2014 for Geneva (MeteoSwiss 2014; NCDC/NOAA 2014), though with significant gaps (about 25% of the TDB data is missing). The gaps do not appear to have a bias, like one particular season that is consistently missing. Pending further investigation, we assume that the measurement errors are uniformly distributed and do not colour the statistical characteristics of the time series. When the data gaps were too large (e.g., a whole month from one year), we removed that entire year. When the gap was of a few hours, we used interpolation

Table 3.1 – *ASHRAE design temperature percentiles for Geneva. All TMY values are taken from the header of the TMY file, except for the 98^{th} percentile. This was calculated, and so represents the 98^{th} percentile of the 'mean' signal.*

Percentile	**Geneva** (°C)				
(%)	Recorded	TMY	Synthetic	RCP4.5	RCP8.5
99.6	31.13	30.05	30.80	32.56	34.43
99.0	29.21	28.33	29.00	30.24	31.93
98.0	27.35	26.80	27.20	28.00	29.56
50.0	10.10	10.00	10.41	9.66	10.53
2.0	-2.77	-3.70	-1.90	-4.85	-4.09
1.0	-4.04	-5.00	-4.80	-6.53	-5.80
0.4	-5.63	-7.20	-6.90	-8.56	-7.82

based on discrete solutions to a *boundary value problem*[15].

3.10.1 Probability Distributions and Dispersion

Tables 3.1 and 3.2 show values corresponding to the various American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) design temperatures. In searching for suitable "important" extremes, we settled on the widely-used ASHRAE design temperatures. The point of showing these values is to verify that the synthetic weather files effectively reproduce the (design) extremes. Figures 3.19a and 3.19b compare the range (maximum, minimum) and means of synthetic, recorded, and typical time series. The maximum values are represented by the 99^{th} percentile for each month, and the minima by the 1^{st} percentiles. Figures 3.19c and 3.19d show the empirical Cumulative Density Functions (eCDFs) and Figures 3.19e and 3.19f show the PDFs respectively.

Figure 3.19a suggests that the probability distributions of the synthetic series should have truncated left tails compared to the original data, and that is exactly what we see in Figures 3.19c and 3.19e. That is to say, the winter extremes are truncated. Figure 3.19e shows that the resampling procedure tends to smooth the unexpected peaks that exist in the PDF of the TMY data (e.g., between about 5°C and 15°C)[16]. On the whole, the approximation is acceptably close to the measured data.

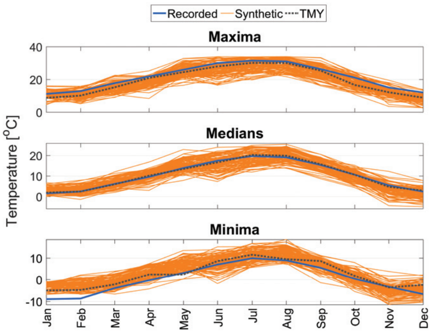
[15] A user-supplied MATLAB function, `inpaint_nans`, solves discrete approximations to one of several partial differential equations, analogous to 'filling holes' in a flat plate. Code downloaded from MathWorks FileExchange. Copyright (c) 2009, John D'Errico.

[16] Unexpected in the sense that said peak is not present in the PDF of the recorded data.

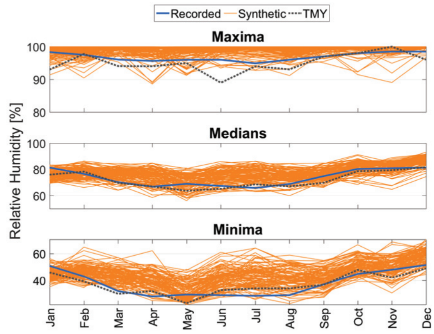
Table 3.2 – *ASHRAE design temperatures for Delhi [left] and New York JFK [right], calculated in the same manner as Table 3.1.*

Perc.	**Delhi** (°C)			**New York** (°C)		
(%)	Recorded	TMY	Synthetic	Recorded	TMY	Synthetic
99.6	42.55	45.00	42.00	33.61	31.00	32.10
99.0	41.17	43.50	40.60	32.02	29.00	30.30
98.0	39.81	42.00	39.30	30.51	27.50	28.70
50.0	26.73	27.00	24.69	13.12	8.00	12.55
2.0	9.60	8.67	9.00	-3.40	-5.00	-5.60
1.0	8.58	7.00	7.30	-5.04	-7.00	-8.20
0.4	7.52	6.00	6.30	-6.92	-10.00	-10.70

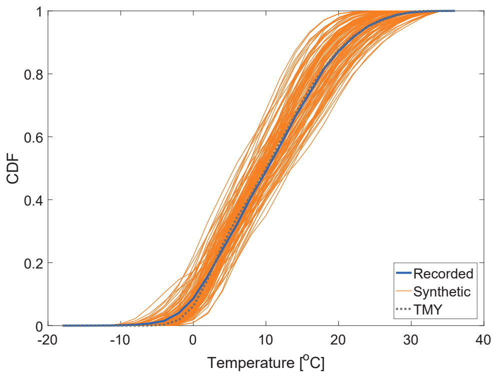
In general, the approach characterises measures of central tendency (mean and median), dispersion (standard deviation and inter-quartile range), and high extremes very well, but is not able to reproduce low extremes (cold temperatures, see the difference in the lowermost lines of Figure 3.19a). That is, the synthetic files contain more extremes in summer than in winter. The cause of this is probably asymmetry in the residuals from fitting models (e.g., Figure 3.12) – if the residuals are skewed, the final generated series will likely be skewed in the same direction. Bootstrapping by itself cannot produce extreme values, since it is merely ‘shuffling’ what already exists. While the values in the TMY source files are recorded values, and may contain extreme values, they are meant to include months whose distribution matches the overall distribution over several decades. This means that the extremes we are looking at are created by a combined effect of bootstrapping and simulation of the SARMA model.



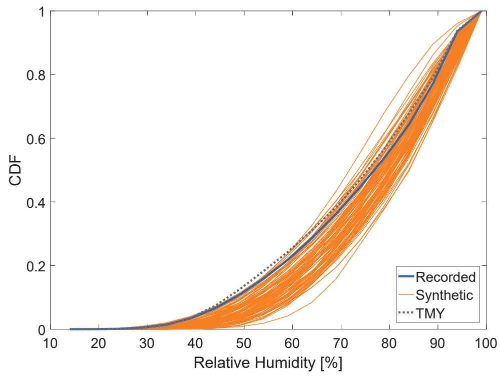
(a) *Extents - TDB*



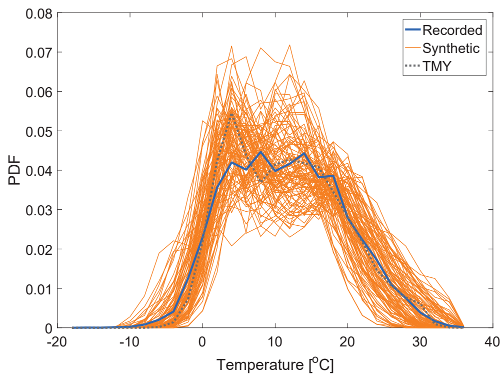
(b) *Extents - RH*



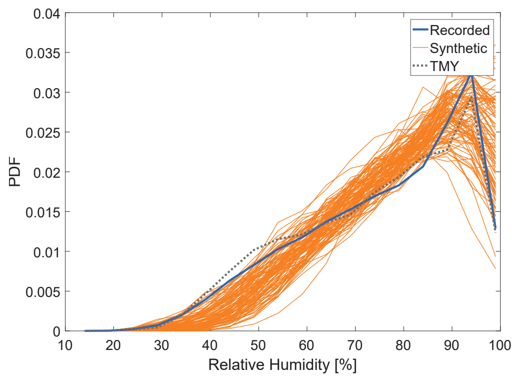
(c) *ECDFs, TDB*



(d) *ECDFs, RH*

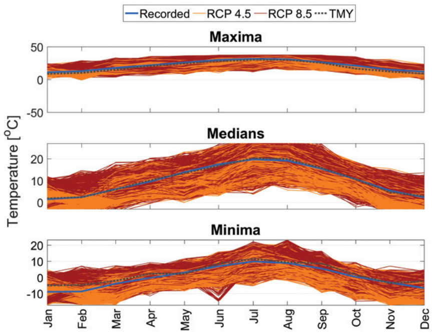


(e) *PDFs, TDB*

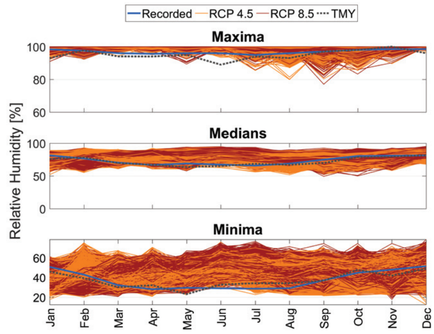


(f) *PDFs, RH*

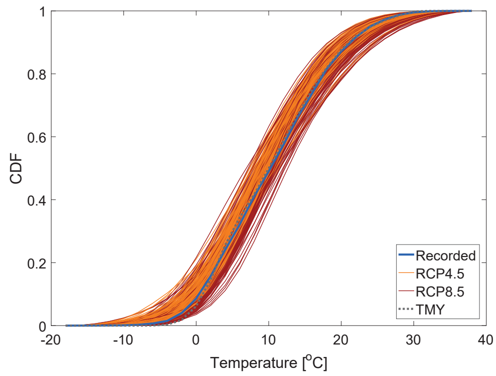
Figure 3.19 – *Monthly extents, eCDFs, and PDFs of measured, typical, and synthetic hourly values of the TDB and RH time series for Geneva. The solid blue line represents measured data, dashed orange represents synthetic, and dotted grey is for TMY data.*



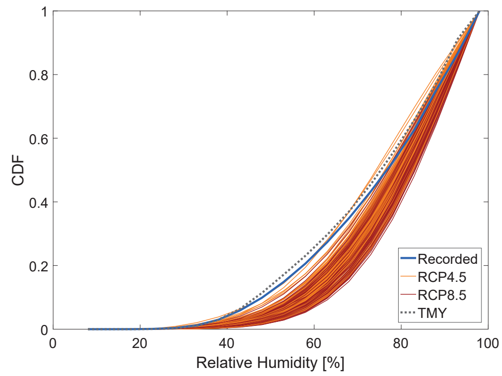
(a) *Extents - TDB*



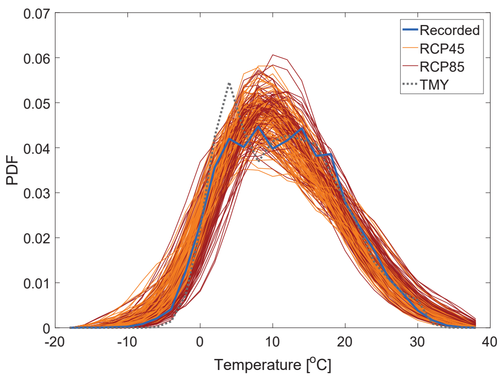
(b) *Extents - RH*



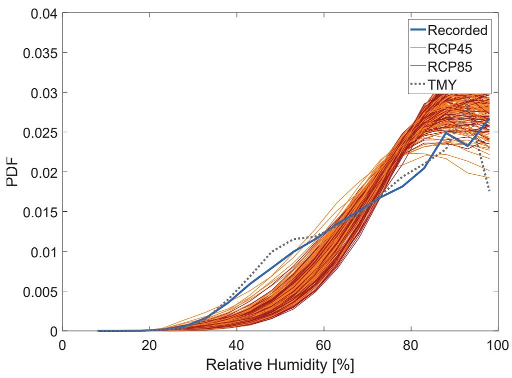
(c) *ECDFs, TDB*



(d) *ECDFs, RH*



(e) *PDFs, TDB*



(f) *PDFs, RH*

Figure 3.20 – *Monthly extents, eCDFs, and PDFs of measured, typical, and synthetic hourly values of the TDB and RH time series for Geneva. The solid blue line represents measured data, dashed orange represents RCP 4.5, dashed red is for RCP 8.5, and dotted grey is for TMY data. The variation in the generated data almost washes out the difference between the two RCPs.*

3.10.2 Sequences

The idea behind checking the presence of sequences above and below some value in the synthetic data is that synthetic series should represent episode intensity and duration (i.e., spell length and magnitude of some temperature) accurately, in addition to the basic descriptive statistics examined above. This is one of the criteria put forward by Lund (1995) to judge the quality of a typical year for building simulation or solar energy studies (Section 2.6.1), though it is highly unlikely that a typical year contains any episodes like heat waves. Hansen and Driscoll (1977) and Magnano, Boland et al. (2008) use thresholds specific to their case studies. Wilcox and Marion (2008), in the creation of TMY files, use the 67th and 33rd percentiles for persistence of warm and cold spells respectively[17]. In search of a sufficiently general set of thresholds, we decided to use the ASHRAE design temperatures that are part of TMY files in addition to the 67th and 33rd percentiles. Looking at the eCDF in Figure 3.22, we see that the synthetic data does include periods of extreme temperature (spells) with roughly the same frequency as recorded data for Geneva. The frequency of occurrence of high temperature episodes, i.e., heat waves, is a little less common. The spell recreation for Delhi (Figure 3.23) shows that the frequency of episodes of all lengths increases slightly, while New York JFK (Figure A.7) shows either slightly higher or slightly lower frequencies. We discuss the recreation of episodes again in Sections 3.12 and 5.3.7, since an issue of concern with the current method is its inefficiency in generating long-duration departures from the mean, like heat waves.

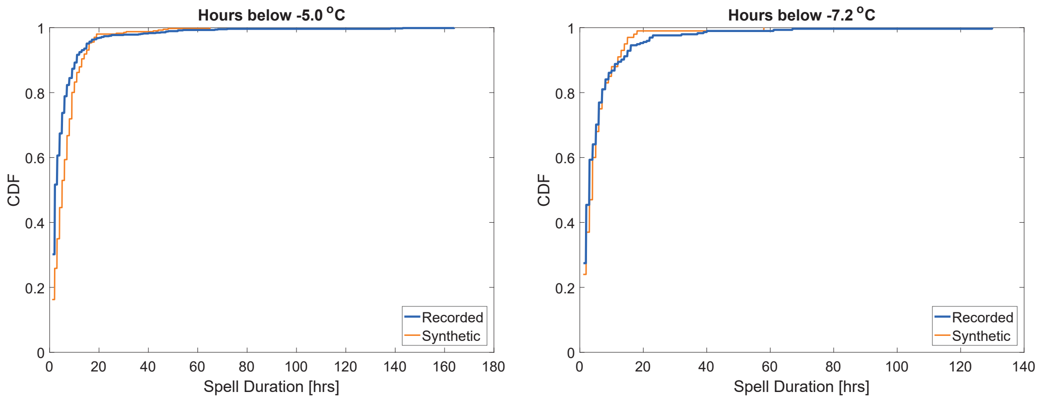


Figure 3.21 – *ECDFs of the low-temperature spell durations in recorded and synthetic data (without climate change forecasts), Geneva. The thresholds are the ASHRAE design temperatures for Geneva, as percentiles: [left] 1.0 and [right] 0.4.*

[17] Note that the most widely-used typical year procedure, the TMY algorithm, deliberately excludes episodes of unusual temperature, i.e., it uses the presence of spells to *disqualify* candidate months(Section 2.6.1).

3.10.3 Cross-Correlation

We are interested in the cross-correlations of temperature with solar radiation and humidity. Unlike vector approaches (e.g., Hong and Jiang 1995; Lee, Sun, Hu et al. 2012), our procedure does not explicitly calculate, or take steps to maintain, cross-correlations. Instead, the resampling procedure for the RH and nearest-neighbour bootstrap for the solar quantities (Section 3.9) preserves the cross-correlations[18].

Table 3.3 – *Correlation coefficients for dry bulb temperature (TDB) with RH and GHI.*

Series	RH		GHI	
	ρ_p	ρ_s	ρ_p	ρ_s
Original	-0.54	-0.5	0.52	0.47
TMY	-0.5	-0.45	0.52	0.44
Synthetic	-0.44	-0.38	0.43	0.36
RCP4.5	-0.3	-0.25	0.35	0.3
RCP8.5	-0.26	-0.22	0.34	0.3

We examine correlations with Pearson's ρ_p and Spearman's ρ_s correlation coefficients. The quantity ρ_p measures the *linear* correlation between two variables, and ρ_s assesses how well the relationship between two variables can be represented as a monotonic function (Dodge 2008). The temperature shows a somewhat linear (negative) correlation with humidity. Mild (positive) correlation is also evident between TDB and GHI. However, this might be a function of the relative sunniness of the example climate. Pending verifications for very cloudy climates, this particular relationship must be treated with caution. The correlations with both RH and GHI are not appreciably different for the synthetic time series.

[18] RH blocks are moved in tandem with the TDB blocks.

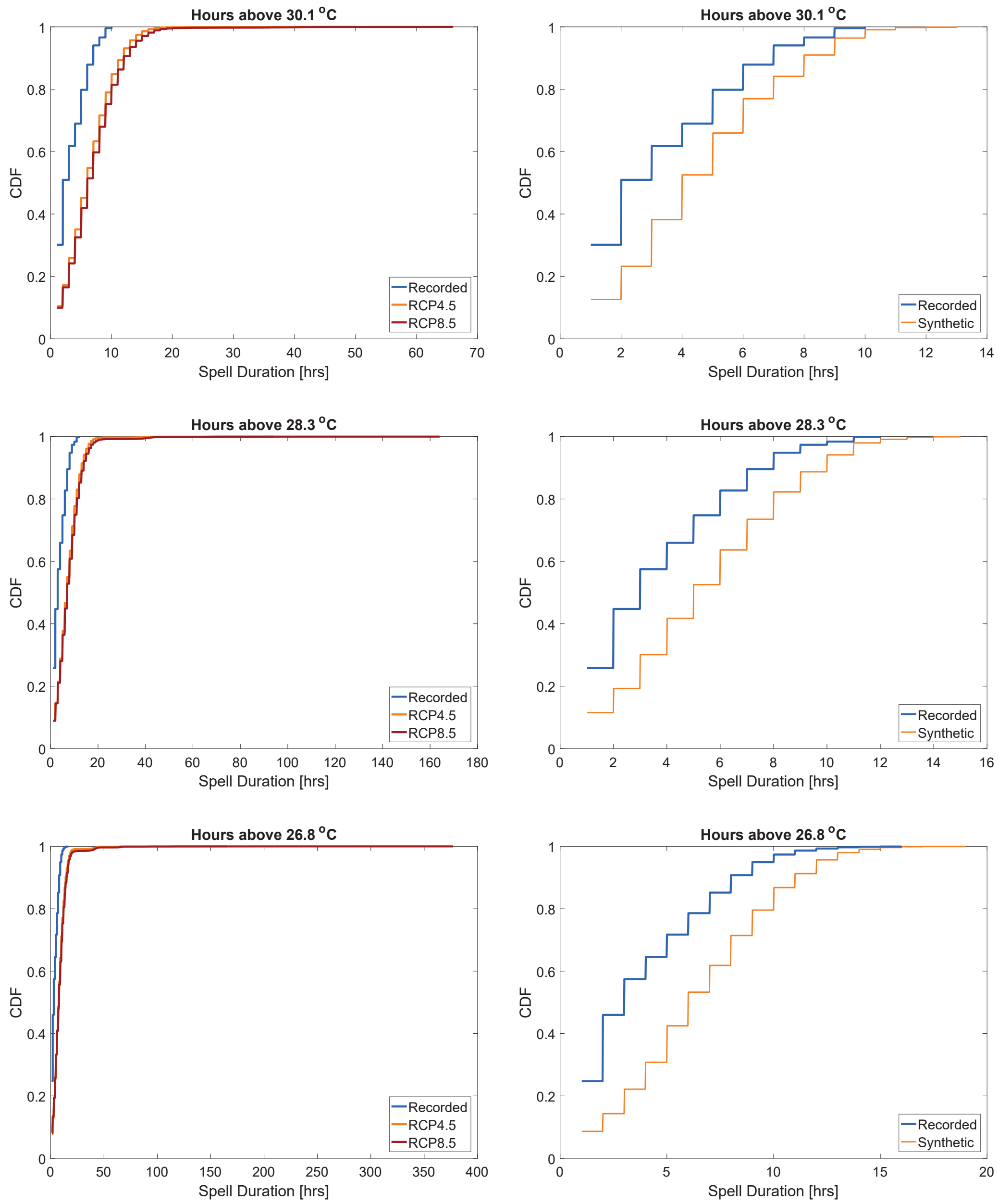


Figure 3.22 – *ECDFs of the high-temperature spell durations in recorded and synthetic data, Geneva. On the left are synthetic files with climate change forecasts, on the right without. The thresholds are the ASHRAE design temperatures for Geneva, as percentiles: [from top] 99.6, 99, and 98.*

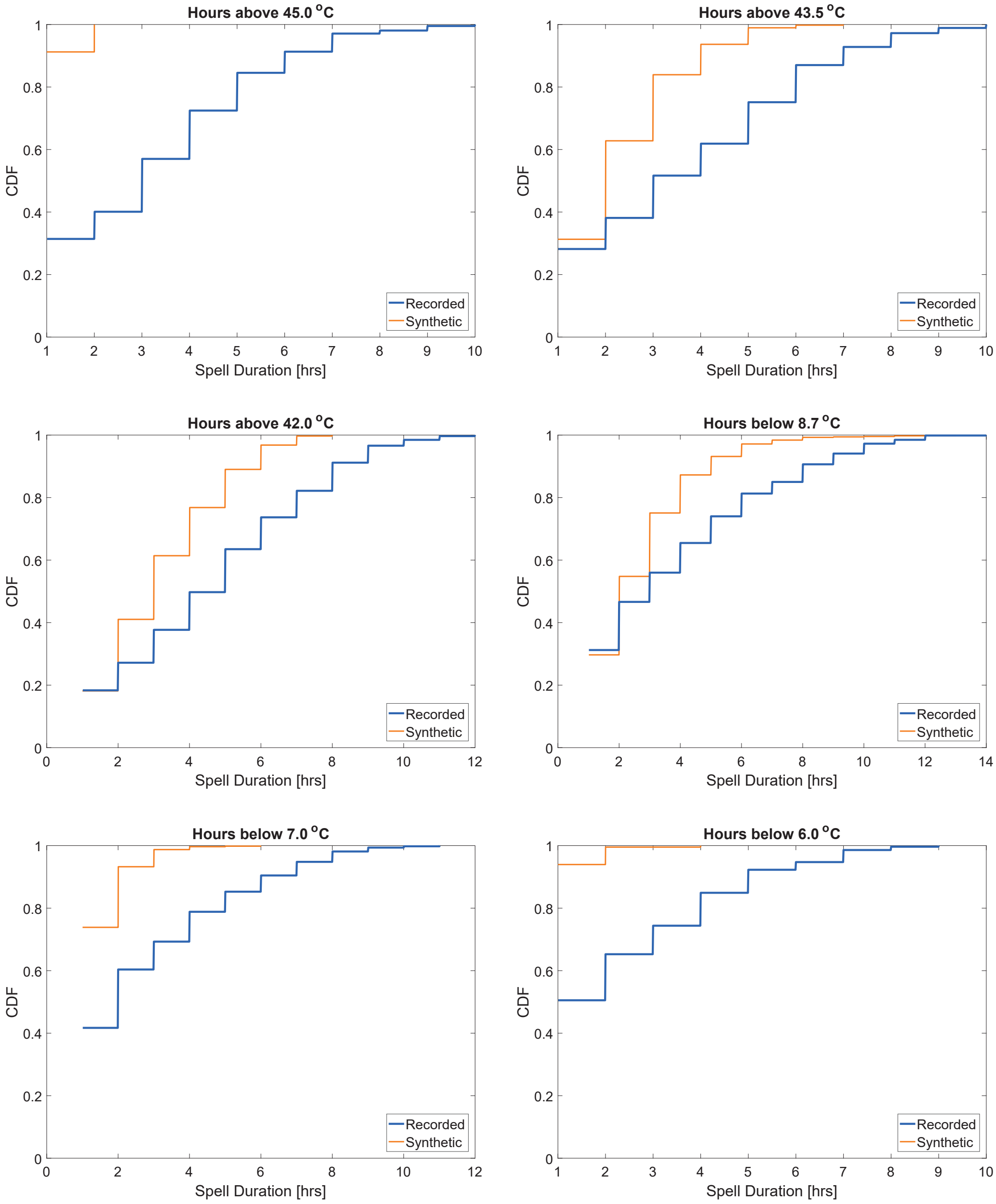


Figure 3.23 – *ECDFs of the spell durations in recorded and synthetic data for Delhi: [from top left] 99.6, 99, 98, 2.0, 1.0, and 0.4 percentiles. Climate change forecasts for Delhi were not used in this thesis.*

3.10.4 Measures of Shape

Skewness and kurtosis are the third and fourth moments of a data set, respectively. Skewness is a measure of the asymmetry of the data. For a distribution to have *skew* is to imply that it is 'heavier' either to the left or right of the mean. Negative values indicate a leftward skew and positive ones a rightward skew. A convenient interpretation of kurtosis is whether a distribution is outlier-prone or not. "Distributions that are more outlier-prone [i.e., sensitive to outliers] than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3" (The MathWorks, Inc. 2015). This is in addition to the interpretation of kurtosis as 'peakiness', where large values above 3 indicate that the distribution is much more 'peaky' than the standard normal. Figure 3.24 shows that there are very small differences (<0.6) between the shape parameters of recorded, synthetic, and TMY series.

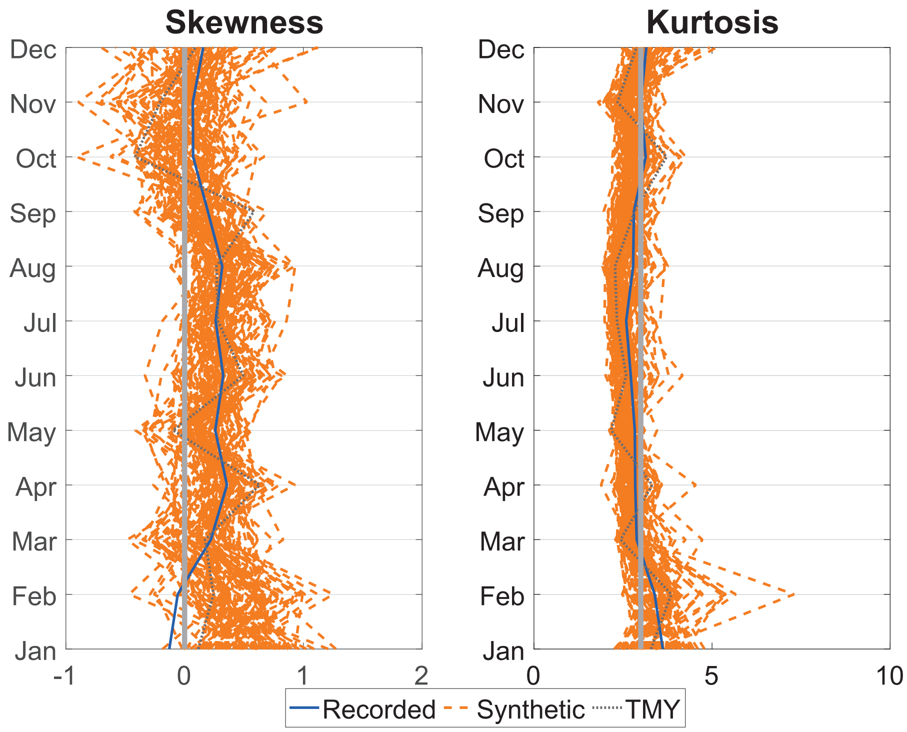


Figure 3.24 – *Recorded data is represented by the dotted line, synthetic by the solid line, and TMY by the dashed line. The thick straight lines in both sub-plots represent the skewness and kurtosis of a normal distribution.*

3.10.5 Random Files with Climate Change

The files generated with an added climate change signal are not necessarily comparable with recorded data. In fact, the point of including the climate change forecast is to say that the weather in the future is *going to be different*. For most of the world, that implies slightly warmer means (1-4°C, depending on which model or RCP one examines). We are not comfortable with a straightforward morphing of current weather by adding a 'mean signal', as discussed in Section 2.2. This means that the weather files we present extend both high and low extremes, implying that we do not necessarily know how the underlying regional climatic systems will change in response to a global rise in temperature. More extreme or unseasonal cold snaps are also possible in the future, like heat waves. Generally, buildings designed for heating-dominated climates will handle colder weather better than a heat wave. A combined plot of 'future' random files for Geneva, compared to the TMY, is given in Figure 3.25. The potential values clearly show larger variation throughout the year than those plotted in Figure 3.16. The time series seem to cut-off at about 35°C. This is a result of the post-processing described in Section 3.8. The maximum temperature in any year can occur between May and September, but does not exceed 35°C.

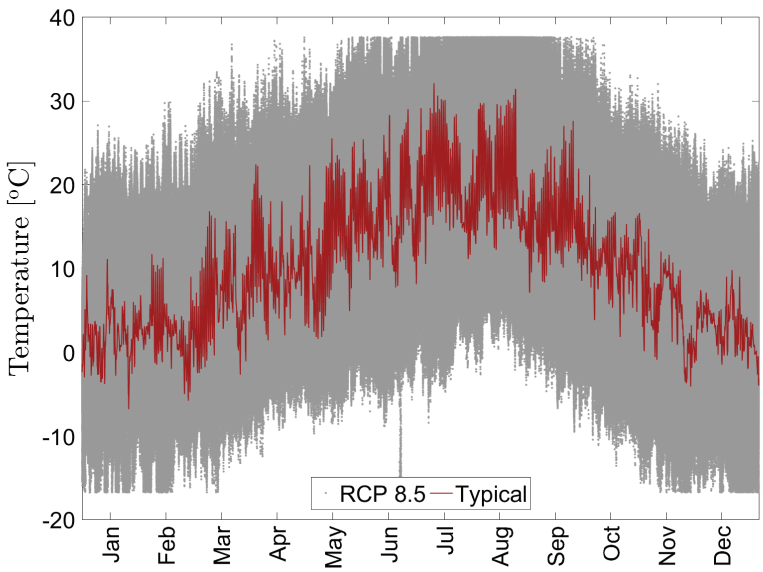


Figure 3.25 – *Synthetic values (RCP 8.5) for TDB, Geneva, plotted with typical values in the foreground.*

3.11 Simulations with Synthetic Files: An Example

Simulations with the plain and future synthetic files described in this chapter are presented in Figures 3.26 and 3.27. In each figure, there are two sub-figures: the top figure (a) shows the simulations separated by date, while the bottom one (b) clubs them by decade to show the distribution (eCDF). In the top figures (a), dots to the left of the line are simulations with recorded data, dots to the right are simulations with future synthetic files using the RCP8.5 projections, and the vertical line is composed of simulations from the plain synthetic files. In the bottom figures (b), the distributions are plotted for each decade, and the darker, thicker, line is the distribution of values from 1981-2013. The year labels in the legend are the starting year of decades[19]. In general, these plots should be treated as what-if or sensitivity analyses. If the user has some way of quantifying confidence in a climate forecast, then these may also be interpreted as uncertainty analyses. Confidence could be quantified by assuming, for example, that the random future climate files generated by our method follow an approximately Normal distribution every year, and that the forecast describes the slow change of mean. The mean also has its own distribution, based on one of the forecasts given by the IPCC (IPCC 2014a) and the CORDEX website (World Climate Research Programme 2015).

In Figure 3.27, the spread of energy use is significant for each individual year, with larger extents for the dominant load – heating. The future values show clear downward (heating) and upward (cooling) trends, respectively. This implies that, while the individual refurbishments will affect the heating load, they will collectively show an upward/downward trend with a warming climate. The trend is less obvious in Figure 3.26, where only the simulations from the base case are plotted (i.e., the home without any refurbishments.). However, the *extents* of the values noticeably decrease and increase for heating and cooling. This is the effect of climate alone, overruling the deliberate variation that we introduced with the refurbishments (Figure 3.27). In Chinazzo, Rastogi et al. (2015b, fig. 12), we showed that the spreads of values for each case increase into the future because the estimates of future temperature diverged. In the simulations presented here, the general warming trend is bigger than the uncertainty surrounding future predictions. In previous work (ibidem, fig. 4), we have suggested weighing future results based on their distance from the present. The method presented in this thesis does not weigh uncertainty differently based on distance from the present. Looking at the procedure in Section 3.7, the same ‘variation’ is added to every year, regardless of our confidence in the values, because we have

[19] Compare these plots to figures 8-12 in Chinazzo, Rastogi et al. (2015b).

no way of quantifying it based on how far in the future it is. This is a possible future expansion of the generating procedure. The eCDFs in Figures 3.26 and 3.27 show that the general shape of the distribution by decade does not alter perceptibly. The upper plots (a) indicate that there should be fewer outliers, which is not noticeable in the eCDFs. The distribution of the past 32 years indicate a mean climate that is somewhat cooler (i.e., lower cooling and higher heating) than future decades.

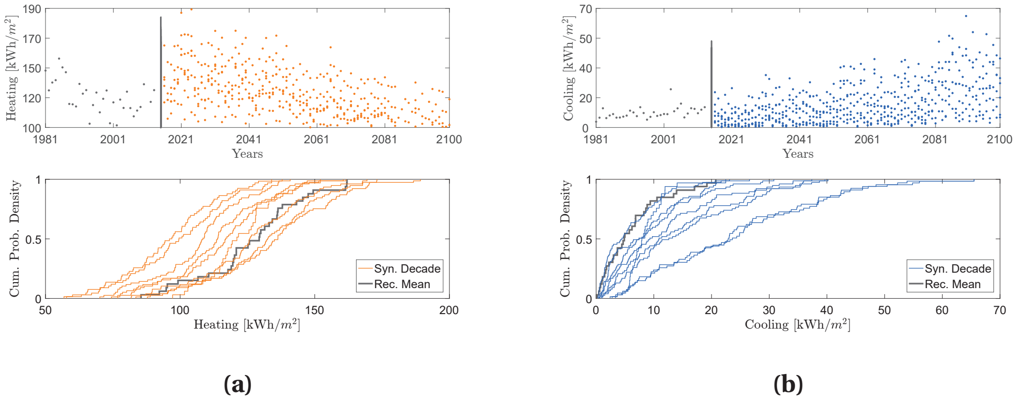


Figure 3.26 – *Simulated heating [a] and cooling [b] energy values for the single-family home in Geneva,* ***without*** *all refurbishment options. The grey dots indicate simulations with recorded data, as do the grey lines in the eCDF plots. The grey lines in the plots on top indicate simulations with the plain synthetic files, which are time-unspecific. The orange dots to the right of the grey line are simulations with synthetic files in which a climate change forecast was included. The synthetic files with climate change always refer to a specific future year.*

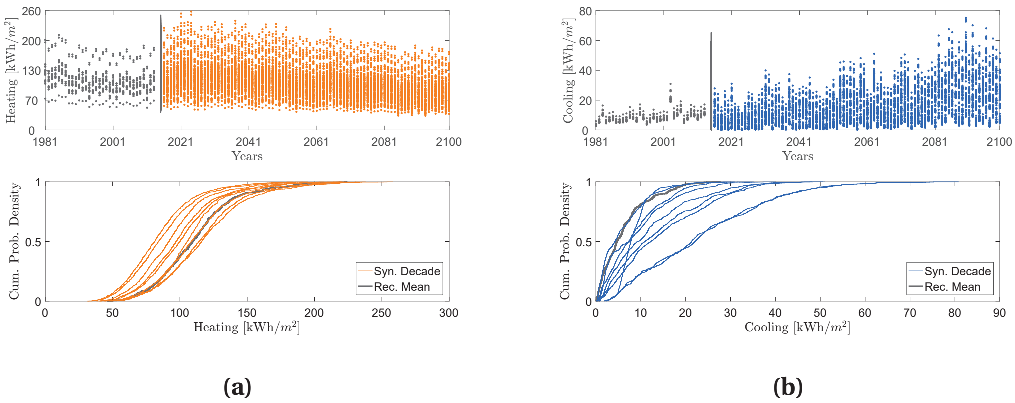


Figure 3.27 – *Simulated heating [a] and cooling [b] energy values for the single-family home in Geneva,* ***with*** *refurbishment options. See explanation of plot details in Figure 3.26.*

3.12 Limitations

3.12.1 Episodes and Extremes on Demand

In our implementation of *stationary* time series models, the generated time series are not meant to allow long excursions from the mean. Since the models are time-invariant, special conditions (or different models) are not automatically invoked to cause a heat wave to persist. Thus any sequences of extreme temperatures are generated purely randomly, and this generator is not efficient for generating heat waves. Extreme temperatures are reproduced well (e.g., Figures 3.19a and 3.20a), they just do not persist for very long. Compare, for example, the length of spells (or persistence) shown in Figures A.7, 3.22 and 3.23. A large number of draws and post-processing might be necessary to obtain synthetic time series with episodes of interest. As we discuss in Section 3.12.4, while the number of runs is independent of the number of weather parameters, a large draw (100-500) may be needed to recreate extremes. It is not necessary to store every synthetic year, of course, and a large pool of files could be built up using years of interest from a set of separate draws.

A possible addition to the weather generator would be the ability to generate episodes, e.g., heat waves, 'on demand'. This could, for example, be done with the use of time-varying models, as opposed to the time-invariant models used here, or long-range dependence models. The reason we eschewed the use of these models so far is that the time-invariant models demonstrated are simpler and have fewer parameters to calculate. Since the simpler answered the needs set out for this project well, parsimony prevailed. The general incapacity of the time-invariant models to produce, on demand, heat waves and cold snaps, i.e., long-duration extreme episodes, necessitates a re-think, if such episodes must be modelled well. The idea behind using time-varying models would be to model two or more 'regimes', defined by the differing number of lags and their coefficients. These regimes could be thought of as representing 'pressure systems'[20] and other influential weather phenomena that cause a particular combination of weather conditions to persist. These systems may occur only a few times a year and therefore have no importance in an annual appraisal of the data, especially one based on a typical year. Recall that the general models by and large include low-order ARMA terms (0-4 lags) and either Seasonal Auto-Regressive (SAR) or SMA terms, or both. Since these coefficients do not change with the seasons, the differences in the stability of summer and winter weather systems (say, a depression) is not modelled explicitly. Yet this persistence may cause a heat wave or similar episode.

[20] "An individual cyclonic-scale feature of atmospheric circulation, commonly used to denote either a high or a low, less frequently a ridge or a trough" (AMS 2015). A low-pressure system is also called a depression.

A low-order ARMA process with high coefficient values, and without seasonal terms, would favour a long departure from the (zero) mean. In future work, we propose to test the use of specialised ARMA models for producing noise series which are able to make a persistent excursion from the mean. This would be a 'heat wave' or 'persistent system' that is applicable only for short periods and not throughout the year. If localised to a specific time, the process could produce an episode without affecting the suitability of the overall fit of the all-season SARMA model. The possibility of transitioning to a persistent regime could be modelled by, for example, a seasonally-varying Markov transition matrix. That is, we strongly suspect that persistent systems tend to cluster seasonally, so the probability of getting one should not be uniform throughout the year. For example, low-pressure systems are far more common during summer in monsoon climates. While this knowledge cannot be acquired from a TMY file, one does not necessarily need to know exactly when such systems occur in a climate, if all one is looking for is to test a heat wave in the summer.

Another class of models are the so-called long-memory models or models incorporating long-range dependence, e.g., those involving fractional differencing ($d \in \mathbb{R}$). Since we dropped integer differencing because it complicated the physical interpretation of the model, fractional differencing did not make the cut either. See Shumway and Stoffer (2011) for a brief introduction to fractional differencing. After some initial tests, we dropped this type of model, but will consider it in future work if it helps to recreate episodes 'on demand'.

3.12.2 Expert Input

A limitation of any method based on fitting conditional mean and variance models to observed data is that the models are *temporary constructs*. These models are based on the characteristics of a sample, not physical equations, so their parameters are recalculated for each new case[21]. We do not claim that these models are better representations of climate than global- and regional-level simulations of physical phenomena. It is incumbent upon the energy modeller who would like to use this strategy to check and recheck the model fits to ensure that crucial assumptions are not violated, and the synthetic time series are physically valid. As we showed in Section 3.5.2, even quantitative selection criteria can sometimes point in the wrong direction.

The choice of re-/sub-sampling for creating synthetic series was motivated by the

[21] Recall the difference between parameters and variables from Section 1.2.3.

general applicability of the method and its relative insensitivity to underlying distributions. A notable set of cases for which the bootstrap is expected to fail is if the data come from a distribution which is in "the domain of attraction of a (non-normal) stable law" (Politis 1998). Since we work with raw material that has a nearly Normal distribution (Figures 3.12 and 3.13), this is not a problem.

3.12.3 Choice of Source File

We do not address the question of data quality separately, folding it instead into the larger discussion of uncertain inputs. The method described in this chapter only partially makes up for the quality of the input/source file. It is up to the user to check the quality and representativeness of the input files being used for a particular design problem. If a file was composed with high-quality data from a long period of record, then the synthetic data based on it is more reliable, and vice-versa. The years that make up the TMY file also make a difference. For example, the years included in the TMY file for Geneva are older (1980s-90s), while the New York stations have newer years (1980s-2000s). In both cases, the recorded data is considerably longer (1950s-60s to 2000s), and the TMY/synthetic files tend to be *warmer*, especially in summer. The effect is also seen in other stations from North America (like Chicago). While we focussed on working with small weather data sets (the typical year files), using more data, particularly recent records, is not detrimental to the generating procedure described here. For example, the approach described in Magnano, Boland et al. (2008) uses eight years of data. We did not focus on this application since we wanted to specifically address the issue of patchy data availability (see section 2.7). In specific applications, we encourage the user to obtain as much recorded data as is feasible for a given site – especially to compare the synthetic data with.

As discussed in Section 2.7, this procedure does not account for spatial uncertainty. If a file is used from a station that is very far from the site of the building, then the difference due to urban and natural factors cannot be compensated by the input modifications proposed in this chapter. The occurrence of extremes and the general shape of the data may be very similar for nearby stations, but persistent micro-climatic effects (e.g., Urban Heat Islands (UHIs)) cannot be accounted for. To examine the stability of the procedure when using different stations in the same city, we looked at New York. The shape and extents of temperature and humidity for the three stations in New York are similar (Figure 3.28). In both cases, the synthetic weather files are warmer than the recorded data, but they are different by about the same amount.

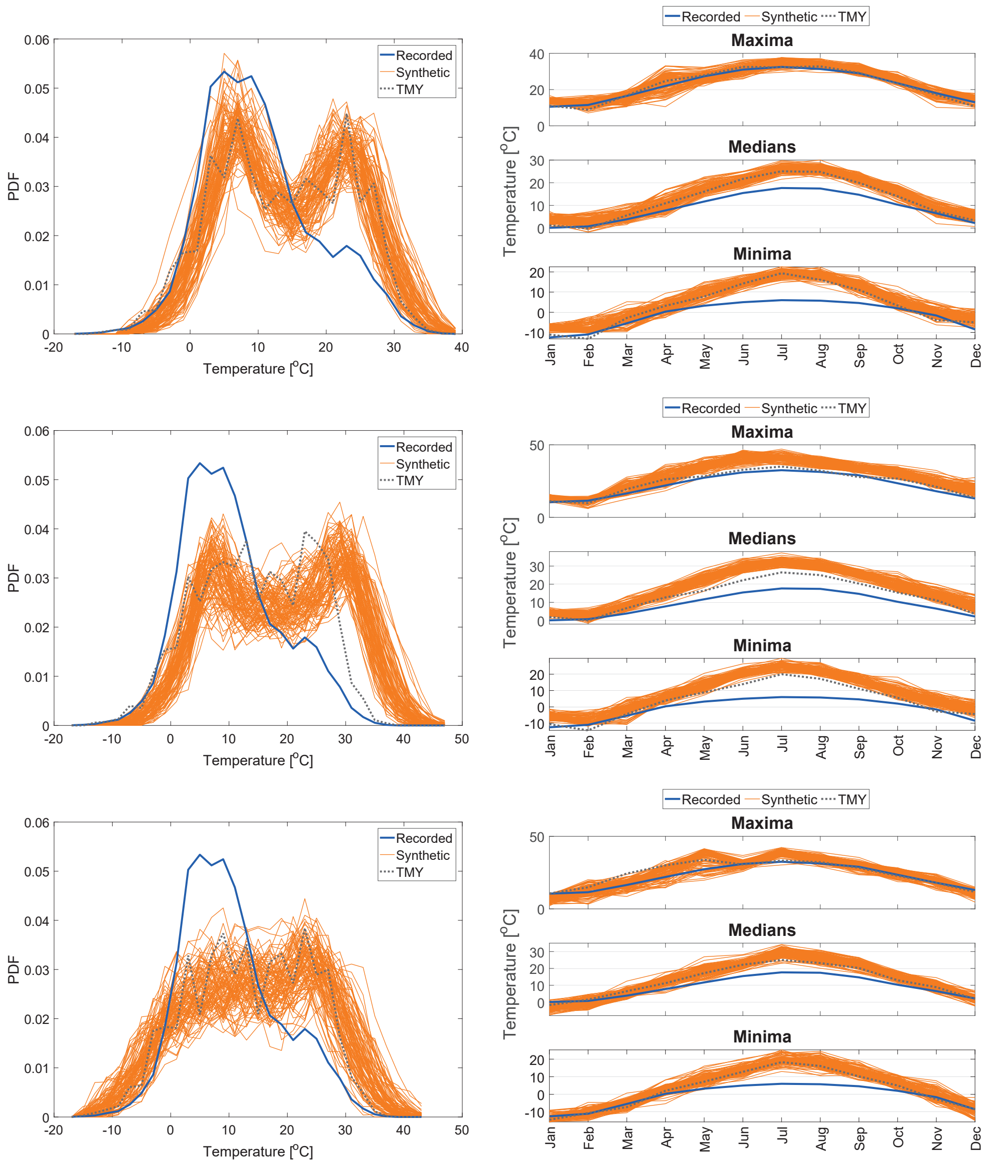


Figure 3.28 – *New York JFK [top], LAG [middle], and CPR [bottom]. The synthetic files follow the distributions of the TMY file closely. The recorded data have a different distribution from the (newer) TMY files, showing a significant 'bump' between 0-10°C. This also pulls the means lower, which is particularly visible in the summer statistics (right).*

3.12.4 Number of Runs

Throughout this procedure, we have worked with either 50 or 100 resampling runs, meaning 50 or 100 SARMA simulations. This was a convenient number of simulations

Table 3.4 – *ASHRAE design temperature percentiles for Geneva using 50 and 100 runs.*

Percentile (%)	**Geneva** (°C)					
	50 runs			100 runs		
	Synthetic	RCP4.5	RCP8.5	Synthetic	RCP4.5	RCP8.5
99.6	30.80	31.43	33.00	30.80	32.57	33.16
99.0	29.00	29.20	30.64	29.00	30.08	30.63
98.0	27.20	27.11	28.42	27.20	27.71	28.24
50.0	10.41	9.62	10.43	10.41	8.97	9.59
2.0	-1.90	-5.27	-4.39	-1.90	-6.05	-5.19
1.0	-4.80	-6.99	-6.14	-4.80	-8.02	-7.17
0.4	-6.90	-8.97	-8.23	-6.90	-10.36	-9.57

in terms of computational time and variation. As with any random number or Monte Carlo (MC) simulations, the number of runs is independent of the number of variables being modified. However, the desire to produce extremes may force the user to run more simulations. Table 3.4 and fig. 3.30 compare results from two run sizes. Predictably, the 50-sample run tended to produce less extreme values than the 100-sample run, and was less consistent in producing these extremes. Since the generating procedure is relatively cheap, we the user may run the generator until some desired extremes or heat waves are reproduced. It could be possible, for example, to run several 50-sample runs and then combine the result. This selective retention of synthetic years will, however, influence the final distributions (i.e., the output will be biased). If one keeps *all* of the generated files from many repetitions, the probability distribution should be stable. However, if one discards similar, moderate, files from each run, then the probability distribution will tend to become heavy-tailed (increased kurtosis). Ultimately, the use of synthetic weather data is a 'brute force' approach to the characterisation of uncertainty due to weather in buildings. That is, the only way to improve one's estimate of bias in some statistical measure of model output is to create more paths (n_{sim} and n_{boot}) and simulate each resulting time series separately. We expect that the time required to improve one's coverage would scale, at best, linearly with the number of simulations, i.e., $O(n)$. That is, 200 simulations should take twice as long as 100 simulations, up to some limit where post-processing the data becomes time-consuming in itself.

The representativeness of the source file influences the representativeness of the synthetic series, and the user may not be comfortable with assuming underlying distributions for future values. If the user is not bound to preserving the probability distribution, then not using the entire set of 100 files from one generation run (or many hundreds of files from several runs) is feasible. It is quite reasonable that several hundred files may be generated until a desired set of characteristics is achieved (say, extreme summers). Then, the user may discard files that are too similar since they will likely produce very similar outcomes. This choice naturally depends on the quantity of interest, and may be made with the regression-based emulator proposed in Chapter 4. As we discussed in Chapter 1, we suggest that using the *posterior*, i.e., the results of an emulator, is better for selecting weather years of interest than a *prior*, i.e., the raw files themselves. In most cases, the point of this selective exercise would be a what-if analysis, e.g., the effect of an extreme heat wave. The selection of weather files will necessarily introduce bias.

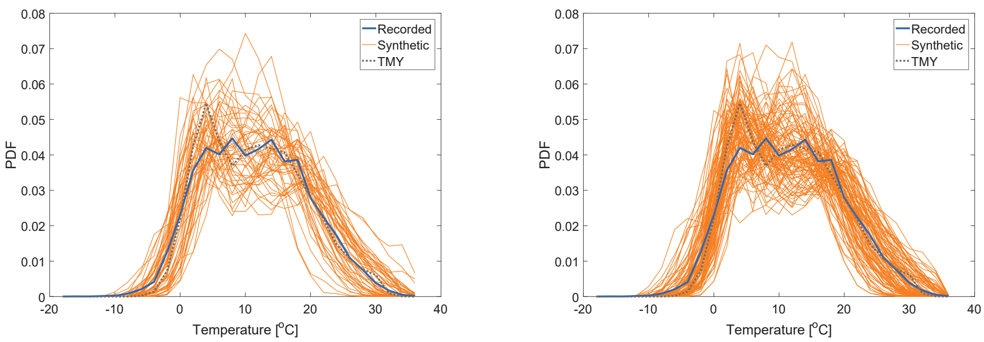


Figure 3.29 – *TDB values for Geneva generated with 50 [left] and 100 [right] runs. The extremes are not visibly different, but they are slightly smaller in the 50-run dataset.*

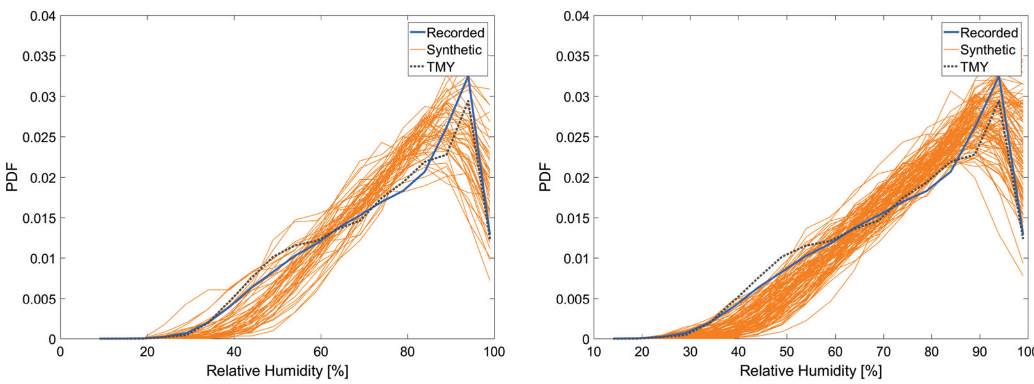


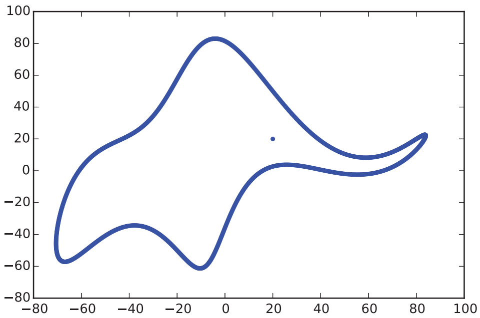
Figure 3.30 – *RH time series for Geneva, using 50 [left] and 100 [right] re-sampling runs.*

3.13 Synthetic Generation: Summary

We have demonstrated the use of Fourier fitting, conditional mean models, and resampling to create synthetic weather data. This synthetic weather 'data', while based on a typical year file, turns out to be more representative of the range of values seen in our example climate over the last sixty or so years than the typical file on which the procedure is based. The descriptive statistics, correlations, and dependence structures of the synthetic time series greatly resemble those of the original source material, the TMY. In addition, the synthetic series also reproduce or exceed extremes in measured data. The innovation of including time series models at multiple scales, i.e., separate models for annual and daily variability, based upon the work of Boland (1995) and Magnano, Boland et al. (2008), is partly responsible for this performance. The mixing of persistence effects at different time scales (e.g., daily and seasonal moving average or auto-regressive structures) is better represented by a model composed of separate parts explicitly representing these diverse time-scales.

The user cannot outsource all judgement to the statistical techniques. Even something as well-understood and pervasive as the Fourier transform still requires expert verification for appropriate usage. Errors and inconsistencies are bound to occur, because the generation procedure does not 'know' which values are unreasonable. Checks for which extreme is too extreme are much more difficult to hard-code into a procedure than physical checks, like removing RH values above 100%. In any case, aggregate quantities like annual sum of energy are less influenced by the occasional 'unrealistic' value, however that is defined, than instantaneous quantities like peak load. In an assessment of a noise-sensitive quantity like peak load, the user is well advised to use censoring of the kind described in Section 3.8.

4 Emulators for Uncertainty and Sensitivity Analyses



With four parameters I can fit an elephant,
and with five,
I can make him wiggle his trunk.

John von Neumann

(Mayer, Khairy et al. 2010)
Image redrawn using a Python script
from Piotr A. Zolnierczuk

4.1 General Approach

In practice, building simulations are *deterministic simulations*: a simulation always gives the same output for a particular combination of inputs, if numerical errors are neglected[1]. The work presented in this thesis is a first step toward changing this practice, i.e., by conducting and interpreting building simulation in a *stochastic paradigm*. In this paradigm, uncertainty or sensitivity would be quantified explicitly, through variability intervals[2], either constructed empirically with Monte Carlo (MC) simulations, such as by using the synthetic inputs from Chapter 3, or through frequentist interpretations of regression predictions, such as in this chapter. With the synthetic weather data, we presented (pseudo-) random inputs that might be used to carry out stochastic building simulation. To enable a practical application of those

[1] Errors arising in finite precision calculations, especially in the solution and approximation of thermal networks and shading calculations.
[2] The extents of variation in outputs like energy use.

computationally-intensive analyses, this chapter presents a choice of rapid-response regression models. These models may also be used for numerical Sensitivity Analysis (SA)[3]. The input variables used in the regression models presented here may be extended or amended to suit the situation (design problem) that a user is working with.

These regression-based emulators or meta-models are meant to be rapid-response approximations of a building simulation, or building system. As explained in Chapters 1 and 2, sensitivity and uncertainty analyses do not themselves *need* emulators. In fact, since these are 'simplifications' or 'surrogates' for full-scale simulation, one ends up measuring the sensitivity of the emulator, or the propagation of uncertainty through it, and not the characteristics of the original system. It is the large number of simulations required by sampling-based methods (see Section 2.4) that force the use of emulators. In Section 2.5, we gave examples of simplified methods and emulators from the literature. Unlike previous efforts, though, and consonant with the philosophy of the chapter on time series models (Chapter 3), we will *not propose a unified model* that works across all climates and buildings. Instead, we propose a structure to *build reliable models* cheaply and on-the-fly, *customised to a design problem.* We will try to convince the reader in subsequent sections that trying to make a unified or general regression model (for all buildings in all climates, or one climate) is both complex and counter-productive, whereas training an emulator for each design problem, i.e., a particular building defined by its geometry and usage in a particular climate, is far more robust.

Since the application of Gaussian Process regression is novel in building simulation, we compare it with classical models. Gaussian Process regression either relaxes or modifies many of the conditions presented in the discussion for classical models (Section 4.3.1). Conditions on the inputs and outputs are most restrictive for the classical Normal model, and the goal of using the other classical models presented here, Generalised Linear Models (GLMs) and Generalised Linear Mixed-Effects Models (GLMMs), is to progressively relax these conditions. We show why the considerably freer Gaussian Process (GP) models are, in the experience of the author, best suited to the problem at hand.

Regression is a key field of interest in statistical or numerical studies because it gives a user the ability to predict the effect of one or more *explanatory* variables on one or more *response* variables, i.e., "what happens to [some response] y as [some input(s)] x

[3]That is, quantifying the sensitivity of output(s) to input(s) through repeated simulation. See Sections 1.5.5 and 2.4 for more details.

varies" (Davison 2003, chpt. 8). Essentially then, regression analysis is the search for some function

$$y = f(x), \tag{4.1}$$

where x is a non-random input (which can be continuous, discrete, etc., and is either controlled by the experimenter or not), y is a random output, and $f(x)$ is some function. The aim of a linear regression exercise "is to disentangle systematic changes in y due to variation in x from the haphazard scatter added by the errors ε" (ibidem, chpts. 8). We will build up the case for our choice of model, Gaussian Process regression, by starting from the simplest possible models – the classical Normal Linear Models (NLMs).

In our application, the response is either annual heating need or annual cooling need, per floor area, also known as the Energy Use Intensities (EUIs) for heating and cooling. In this thesis, the annual heating/cooling need or intensity is defined as the annual sum of energy used for space heating or cooling divided by the floor area. We will generally use the terms 'energy need' or 'energy use' with units kWh/m^2. This quantity is equivalent to stating the energy used in kWh, i.e., not divided by floor area, but the normalizing step makes the energy figures comparable across different buildings. Other quantities of interest in simulation include the (instantaneous) demand, in W (Watts). The building 'loads' discussed in this thesis do not account for the inefficiencies of Heating, Ventilation, and Air Conditioning (HVAC) systems, or the electric/gas grid.

This chapter deals first with the inputs, then with models, and finally diagnostics. The results from fitting different types of models are of two types: a 'first' model, trained only on typical year files; and a 'best' model, trained on typical, recorded, and in some cases synthetic, weather data. The best fits for each kind of model are presented alongside the description of that model, and the reasoning behind selecting them is discussed in Section 4.2.6.

4.2 Inputs and Outputs

4.2.1 Case Studies

This chapter discusses the procedures for developing a classical regression model for two kinds of sensitivity analyses explored through two case studies – a restricted analysis and a comparatively general one. In this exercise, we have access to a large amount of simulated data for training and testing. The restricted analysis (Case 1) examines the sensitivity of one building, a single-family home, to changing weather conditions in one climate (Geneva). Additional variation is introduced by refurbishments to the envelope (Table B.4). The general sensitivity analysis (Case 2) also examines the effect of changing weather conditions, but this time for several different building types *together*, with some variability in their material properties (expressed through their 'age' or 'era', details in Table B.3). These variabilities may also be interpreted as uncertainty in the envelope or building properties, which interact in complex ways with weather variability. The building types differ in their usage as well as in their geometry. Comparing the results of the restricted and general sensitivity analysis contrasts the decision to use a unified model (i.e., a single regression model for any building type) against a model specific to a particular building (geometry/usage). The examples discussed in this chapter are theoretical, since they are not presented as a particular design problem, where one would be dealing with only one building in a particular climate.

An overview of the case studies is presented in Figure 4.1, and details of the simulations are in Section B.4. The general, cross-building, sensitivity study is carried out on the DOE Commercial Buildings Reference Database (Deru, Field et al. 2011), hereafter referred to as the United States Department of Energy (USDOE) buildings. The USDOE database has sixteen distinct sub-types/sub-blocks of buildings, grouped into eight overall categories based on usage (Block/Type in Figure 4.1 and table B.3). Each sub-type in turn has three variations for envelope construction, which are representative of an 'era': pre-1980, post-1980, and new construction. Since these are all commercial buildings (except for one mid-rise apartment block), we added a typical single-family European house to the mix (the first case study). The base case construction of the home in Geneva corresponds to typical practice from the Italian residential sector in the 1970s, before the advent of energy or thermal regulations. This example is an extension of the case study analysed in Chinazzo (2014), Chinazzo, Rastogi et al. (2015a,b) and Rastogi, Horn et al. (2013). The USDOE buildings are abstract representations of a handful of commercially significant usage types from the United States, not actual buildings. Each individual USDOE building (and the home) represents an experimental unit, but we simplified the experiment by considering only distinct usage types as units. So, for example, the three sizes of offices are combined into one category – office (Figure 4.1).

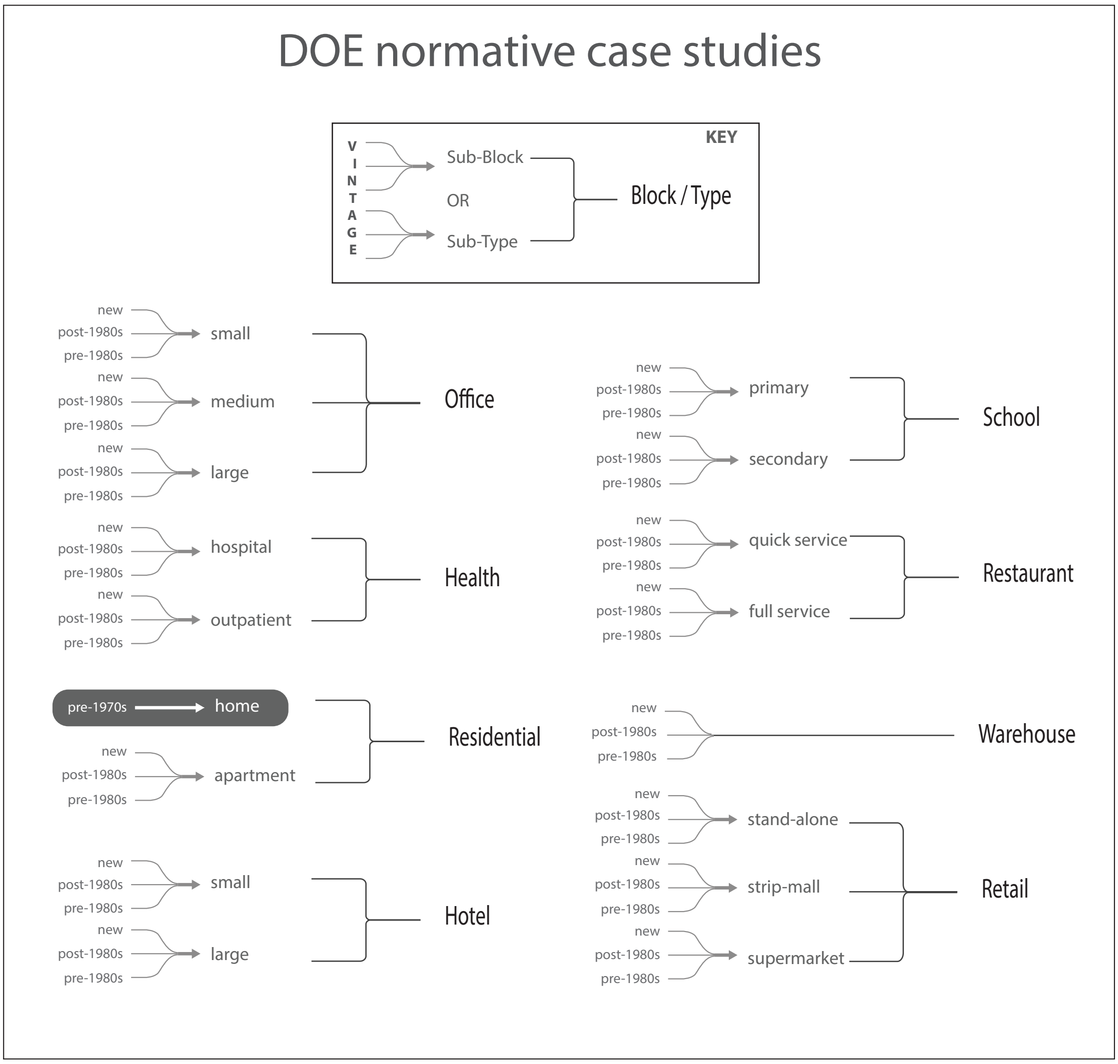


Figure 4.1 – *The case studies drawn from the United States Department of Energy (USDOE) Commercial Reference Buildings Database (Deru, Field et al. 2011). Details are in Table B.3.*

4.2.2 Choosing Inputs: Orthogonality and Dependence

Using an emulator for building simulation entails a loss of information (assuming that the simulation is the ground truth), so we aim to use the least number of predictors or independent variables *without losing too much information*. Throwing out some candidate predictors is known as *dimensionality reduction*, helping to create parsimonious emulators. The inputs, predictor variables, or explanatory variables for this regression exercise are an assortment of climate and building properties[4]. A

[4] 'Predictors' are often called 'explanatory variables', 'independent variables', or 'covariates'.

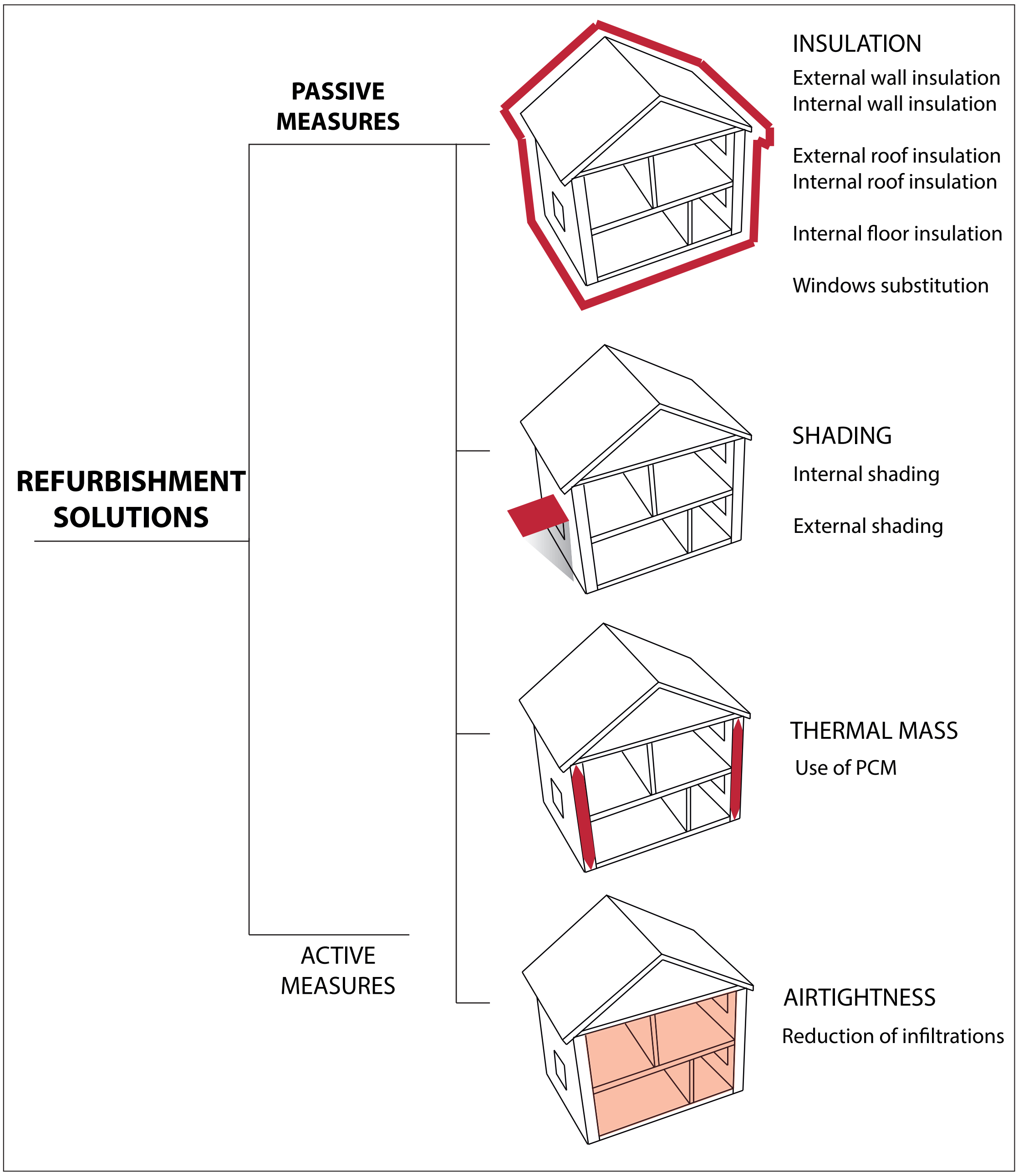


Figure 4.2 – *Refurbishments for the home in Geneva. Details are in Table B.4.*

large number of related initial candidate predictors were considered[5], classified by their source: building-related, climate-related, and mixed. A procedure based on correlation estimates and Principal Component Analysis (PCA) was used to whittle

[5]See Tables B.1 and B.2 for the full list.

the number of inputs down to an approximately orthogonal set of predictors drawn from the larger pool of candidates[6]. The strength of correlation (Pearson's correlation coefficient) between the inputs is shown in Figure 4.3, labelled using the input codes presented in Table 4.1[7]. The choice of regression predictors is governed by the variable building design parameters which are of interest in a particular simulation experiment. For example, if we examine the sensitivity of one building with a fixed geometry, then predictors related to the building geometry (which is a building design parameter that can, theoretically, be varied) will be constant. Therefore, these factors will be taken out of the regression model. That is, if the range of variation of some input, r_{p_i}, is less than some appropriately small cut-off for that variable, ϵ, then that particular input is no longer considered ($p_i \notin P$), and the dimension of the regression space is now $d-1$. This cut-off value depends on the units of the predictors, though in this chapter we convert all inputs to z-scores, making them dimensionless and of comparable magnitude.

Neither the use of correlation nor PCA should be treated as hard rules. Rather, the results of the tests have to be examined in the light of physical knowledge about the system. For example, while Window-to-Floor Ratio and Window-to-Wall Ratio show a significant correlation in our case studies, we know that they describe sufficiently different aspects of a building to merit individual attention. In PCA, the coefficient of each original variable in composing a Principal Component (PC) is indicative of the contribution of the original variable to the variance in the initial basis space of inputs. A coefficient close to zero indicates a small contribution, while one approaching ± 1 indicates a large positive/negative contribution. If two variables are correlated/collinear, then their coefficients will be similar enough to merit discarding one of the initial variables. PCA also offers the possibility of creating a new basis space where the PCs are the basis vectors. However, this would certainly complicate the interpretation of the regression models, since it is difficult to grasp what a basis vector composed of distinct physical quantities, say, Dry Bulb Temperature (TDB) and Dew Point Temperature (TDP), would mean.

The list of inputs should be treated as incomplete. In fact, the concept of a complete list of predictors is essentially meaningless, for two reasons. Firstly, the number of inputs to building simulation is vast, and a regression model could conceivably include any of them. A model that includes all of them would lose its competitive advantage

[6] See Section B.3.2 for a discussion of the orthogonality of predictors, Principal Component Analysis, and why it is better to use as few predictors as possible. Parsimony is generally preferable to avoid unnecessarily complicated models, as discussed in Chapter 3.

[7] The correlation plots of the initial groups are in Section B.3, Figures B.17 to B.19, with a full list of candidate inputs in Tables B.1 and B.2.

over simulation in terms of computational effort. Secondly, the list should reflect inputs of interest in a sensitivity analysis, rather than a large list of predictors that may not all be relevant or orthogonal in the given context.

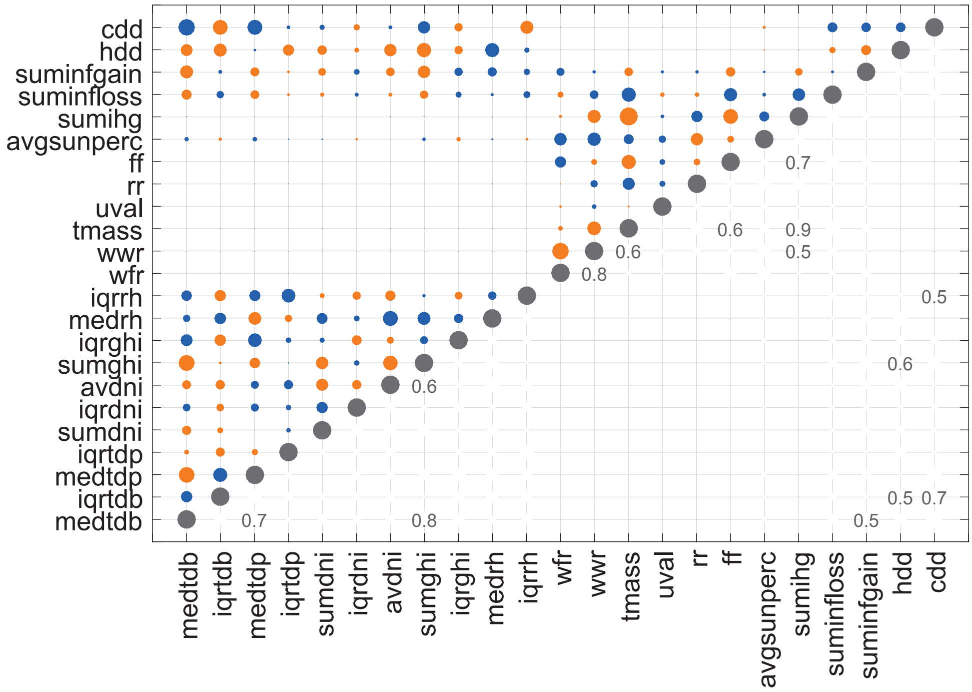


Figure 4.3 – *Correlation coefficients for the final selection of predictor variables, using data from Case 2. See Table 4.1 for the list of codes. Orange dots indicate positive correlation, while blue denotes negative correlation. The size of each dot indicates the strength of correlation (larger dot equals stronger correlation). The black dots on the diagonal are to indicate correlation of 1. The numbers below the diagonal indicate the correlation coefficient (Pearson's) for those predictor pairs where the value was greater than 0.5. Results were similar for Case 1 data.*

The climate-based inputs are calculated from an arbitrary weather file for a given climate region or city of interest. They have nothing to do with the building that is being simulated in that climate. These variables are estimates of population parameters, like the average annual TDB of a climate, calculated from a small sample, e.g., a typical year of data, short records, or synthetic weather years. The representativeness of a descriptive statistic calculated from a single year of data is a matter of contention (see Section 2.6), but this is not an issue in this regression exercise. That is, any simulation may be used as a data point to train a regression model. If the training data set is not typical or representative of the full range of conditions that is of interest to the user, then the regression model will extrapolate badly. Appropriate sampling is discussed more broadly in, for example, Fürbringer and Roulet (1995) and Kleijnen (2008).

The building-based inputs are physical properties, generally of the building envelope, that were selected based on their expected influence on space heating and cooling demand. Building variables are calculated from the building's geometry, material properties, and other inputs without any consideration of a given climate. The building properties are all envelope properties except Thermal Mass, which can be placed both in the envelope and inside the building. See Sections 1.5 and 2.4 for a discussion about why calculated physical properties should also be treated as sample-based estimators of some population parameter of interest from an unknown underlying quantity or its distribution[8]. The building-based inputs may be obtained either from simulation or calculated separately. In the data presented here, for example in Figure 4.3, the values of the input (independent) variables were extracted from EnergyPlus outputs. When considering a new design, these inputs are speculative. In this case, we recommend using ranges of possible values for all inputs, even if variation due to that particular input is not of interest. This will ensure that future changes in that property, either deliberately or through mistakes, are accounted for. In the case of renovations or studies with existing buildings (like the assessment of overheating risk), the user will have to undertake the same forensic effort that is required to create a representative energy model – correct construction details, material properties, etc.

The mixed variables are functions of the interaction between climate and buildings. Though Internal Heat Gain is not a function of the building construction, climate, or their interaction, the *sum of internal heat gain* is included in the regression exercise because it is essential in characterising thermal simulation (see Section 5.3.8 for how better occupancy and equipment modelling can be incorporated later).

4.2.3 Outputs: Distributions and Smoothness

The first case study is a home in Geneva with twenty-four variants based on retrofit options for the envelope and infiltration, and the second case study consists of forty-eight commercial buildings, simulated in a variety of world climates (see Tables B.5 to B.8 and figs. B.22 to B.24). The empirical distributions of simulation outputs[9] from the two simulated case studies are given in Figures 4.4a and 4.4b. The different refurbishment/design options and weather file types have been amalgamated to show

[8] That is, the calculated property is nominal, and the value for a constructed building may be different due to calculation simplifications, construction errors, etc.

[9] Energy use obtained from simulation, the dependent variable.

Table 4.1 – *Final list of predictors for regression. IQR stands for Inter-Quartile Range (IQR). All descriptive statistics are annual, e.g., annual sum. See Tables B.1 and B.2 for details.*

Group	Quantity	Statistic	Code	Units
CLIMATE	Degree Days	Cooling	*cdd*	°C-day
		Heating	*hdd*	
	Dry Bulb Temperature	Median	*medtdb*	°C
		IQR	*iqrtdb*	
	Dew Point Temperature	Median	*medtdb*	°C
		IQR	*iqrtdb*	
	Global Horizontal Irradiation	Median	*sumghi*	$\mathrm{MWh/m^2}$
		IQR	*iqrghi*	
	Direct Normal Irradiation	Average	*avgdni*	$\mathrm{MWh/m^2}$
		Sum	*sumdni*	
		IQR	*iqrdni*	
	Humidity	Median	*medrh*	%
BUILDING	U-value	Average	*uval*	$\mathrm{W/m^2K}$
	Thermal Mass	Sum	*tmass*	MWh/K
	Envelope Ratios	Win-to-Wall	*WWR*	—
		Win-to-Floor	*WFR*	
	Massing	Form Factor	*ff*	—
		Roof Ratio	*rr*	
MIXED	Shading	Average	*avgsunperc*	%
	Infiltration	Sum	*suminfgain*	GWh
			suminfloss	
OTHER		Sum	*sumIHG*	GWh

the distribution of all simulation results. In the first case study, of a house in Geneva (Figure 4.4a), the distribution of heating values looks closer to a Normal distribution than the cooling values, which seem to follow an exponential distribution. The cooling distribution is easily explained: in a mild heating-dominated central European climate, the number of years with lower values of cooling loads *should* be much higher than years with high cooling loads. In Figure 4.4b, neither heating nor cooling values approximate a Normal distribution. The preponderance of small heating and cooling energy use values is likely due to a tendency on our part to pick relatively mild climates, and not because building simulation inherently produces exponentially-distributed output. Individual climates, like in Figure 4.4a, tend to produce distributions that are easier to interpret. In this case, we expect that Geneva will have more years with mild cooling needs, and several years with nearly zero cooling. Heating is dominant in this climate, and will always stay above a certain value, even in the mildest winters.

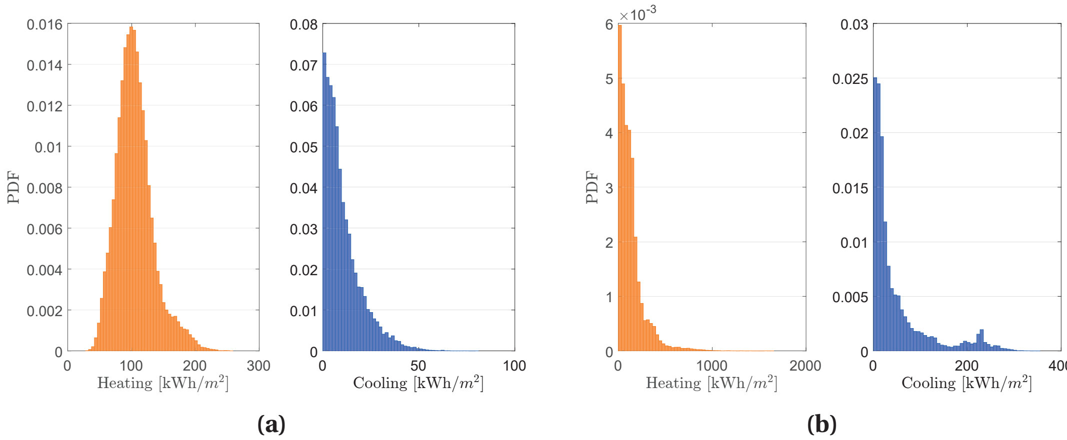


Figure 4.4 – *Histograms of the response variables (y) for the home – Case 1 [a], and USDOE buildings – Case 2 [b]. In each sub-figure, y_{heat} is on the left and y_{cool} on the right. The plots include simulations using several hundred weather files and all refurbishments (Case 1), or all building types (Case 2).*

The data generating process we are looking at, building simulation, is highly non-linear and non-smooth for most dimensions/predictors. There is no guarantee, for example, that a building with twice as much insulation as another will use about half as much energy, even in the same climate. Using an aggregate quantity like the annual sum of energy need smooths the response somewhat, since a summation is not a point estimate of a time series (e.g., peak demand) but its integrated value (i.e., energy used over a year). If substantially different geometries and usage types of buildings (say, the USDOE buildings) are considered together as a super-set, new sources of variation make the response non-smooth again. In simulated building performance data, there is no guarantee of smoothness of response along any 'building' dimension, i.e., by variation of any building property. We found that the response along a 'climate' dimension is reasonably smooth, for a given combination of building properties. Smoothness is an issue in building simulation data since a building's responses follow different underlying models, or 'regimes', in 'regions' defined by the physical properties of the building. These underlying responses need not be so different that they cannot be described by the same class of models, but are unlikely to conform well to a global model. The predictions from classical models (e.g., NLM) were reasonable for the restricted case study, because the building properties did not change much and the sensitivity was explored along dimensions/variables we expect to be smooth, i.e., climatic variables like temperature and solar radiation.

4.2.4 Transformation and Scaling of Data

A common technique to achieve Normality and other desirable properties (orthogonality, matching magnitudes, etc.) of inputs and outputs is transformation and/or scaling. See Section B.3.3 for a brief discussion of scaling/transformation techniques. Given that we are dealing with a space of about twenty predictor variables (covariates), it is more confusing to talk about 'distances' between points if the data has been transformed. We prefer to keep the original basis as far as possible to enable an examination of the data with physical intuition. For example, the 'distance' between points with U-values $U_1 = 1\frac{W}{m^2K}$ and $U_2 = 2\frac{W}{m^2K}$ is half that of the 'distance' between U_1 and $U_3 = 5\frac{W}{m^2K}$. However, this preference for the original units is untenable for two reasons: the vastly different magnitudes of the predictors (often caused by differing units), and the strong *collinearity* of the building-related predictors, which cannot be thrown out for the reasons mentioned in Section 4.2.2. In addition, a scaling and shifting of inputs to have roughly equal magnitudes improves the assessment of the relative impact of each predictor[10]. To avoid a global transform of the data, we use the familiar z-scores

$$z_i = \frac{x_i - \bar{x}}{s}, \tag{4.2}$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation. The z-scores are calculated for each covariate separately, and the resulting transformed covariates approximate the distributions of the original (parent) covariate with zero mean and unit standard deviation, i.e., $y \sim f(\bar{x}, s^2) \to z \sim f(0,1)$. This transformation is often called 'standardisation', and is an 'affine transform' (the quantities are merely shifted, not stretched). The z-scores are unitless.

[10]Which is the reasoning behind the Standardised Regression Coefficient method of sensitivity analysis discussed in Section 2.4.

4.2.5 The Fitting Procedure

The fitting procedure is divided into three steps: first the available data is split into training and testing data sets, then the model is fitted to the training data, and finally the model predictions (of the independent variable) are examined for the testing data (against 'true' independent variable). In the case of Gaussian Process regression, the optimal hyper-parameters of the models are selected using k-fold cross-validation (see Section 4.4.1.2). This was found to be a better estimating method than Maximum Likelihood Estimation (MLE), since the latter tended to overfit the GP model to the training data.

For any exercise in regression, it is good practice to split one's data into training and testing sets. Conventionally, one runs an experiment a certain number of times (say, a factorial experiment) and fits a model to it. Then, one could repeat the experiment to check the validity of the fit that has just been carried out. Predicting on input values within the range of training data is *interpolation*, while predicting outside the training range is *extrapolation*. Since this thesis demonstrates the use of regression to supplement building simulation, the split of training and testing data is generally on the basis of what information may be easily available to a designer. For example, we expect that the designer has some typical weather files to simulate (e.g., International Weather for Energy Calculations (IWEC), Typical Meteorological Year (TMY), METEONORM (MN)), along with ideas about the different kinds of renovations they might consider. This would be the typical 'training' data of choice (labelled 'Typical Fit'). In turn, the testing data could come from simulations with recorded and synthetic weather data years.

For the thesis we simulated far more data than would normally be available in a design work-flow (in excess of 40000 simulations per case study), so the 'testing' data set is usually much larger than the 'training' data. Three types of regression models are presented, distinguished by the type of weather data files used for training (i.e., the sampling strategy to obtain training data): (1) only typical weather data (Typical Fit); (2) typical and recorded/measured weather data (Second Fit); and (3) random selection of weather data (Best Fit). The first sampling strategy represents a scenario where only the typical files are available for training, which is 'business-as-usual' for energy-based design exercises. The second sampling strategy represents a case where the synthetic files described in this thesis are not used. The third describes a case where all three types of weather files (typical, recorded, and synthetic) are used. In the final strategy, training data is increased progressively in 200 steps. The 'best' emulator of the 200 is chosen based on minimising the Root Mean Square Error (RMSE) between regression predictions and corresponding 'true' values from simulation[11]

In the results shown in Sections 4.3.1 and 4.4.1, the reader will notice that using just typical weather data by itself is nearly useless for predictions of building response to a range of weather phenomena. While we have argued that typical data is not a good enough representation of a significant proportion of possible weather scenarios, even in a restricted sensitivity exercise, the strength of this formal demonstration is still surprising. In this chapter, we show the improvement (or lack thereof) of models fit with progressively more training data via error plots like Figure 4.26 (performance compared to the testing data set). In each sampling strategy, the data points to be added to the training set in each iteration are chosen randomly (from the type of weather data applicable). In the 'typical' strategy, there are very few typical year weather files available for most climates so the sampling is severely limited[12]. So much so that in Case 1, where several building factors are constant (because the building itself is the same), there is not enough variation in the typical files to allow the GLMM to converge[13]. In the second and third routines, the choice is large and the results are consequently better. The completely unstructured sampling strategy of the third step may produce inconsistent results, so could be improved with a k-fold cross-validation or similar internal validation step. However, training with recorded and typical data is also a 'random selection' from the full ensemble of possible weather conditions. As we have argued before (Section 2.6), the typical weather file should represent the 'mean' weather, though whether a file based on historical data represents the future mean climate is doubtful. In addition, the recorded data is a restricted, unsystematic, sample. The inputs and outputs used in the GP models are the same as those in the classical regression models. A selection of results is presented next.

4.2.6 Prediction Error Analysis and Model Selection

The idea of emulator selection, like in Chapter 3, should not be taken as a well-posed optimisation problem. Even if we assume that there is some theoretical optimal model for a given physical system, no model can realistically be *optimal* for all conditions on all data sets. In practice, it is quite likely that there will be a cluster of models which perform well for a given data set and are statistically indistinguishable. These should

[11] That is, the root mean square error of a regression model fit, on the corresponding testing data. See, e.g., Figures 4.25a and 4.25b.

[12] This varies between 2 files for Geneva and 6-9 for locations like New York City, because of a multiplicity of stations and sources. See Section 3.12.3 and our previous work (Chinazzo 2014; Chinazzo, Rastogi et al. 2015a,b; Rastogi, Horn et al. 2013) for a discussion of weather sources and differences between stations in the same area.

[13] There was no advantage to assuming non-Normal distributions for outputs at a given set of predictors, so the GLMM is fitted with a normal distribution and identity link function.

be clustered around this mythical optimal model. Prediction error analysis is generally carried out to verify the fit of a model. It is more qualitative than quantitative, since we inspect the shapes and spreads of errors rather than some arbitrary cut-offs for 'good' or 'bad' values. The errors are given by

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}}, \tag{4.3}$$

where $\hat{\boldsymbol{y}}$ is the $n \times 1$ vector of predictions and $\boldsymbol{y}$ are the test sample values. These may be thought of as 'raw' errors. Another type of errors commonly used in prediction error analysis are the standardised errors

$$\varepsilon_{st,i} = \frac{y_i - \hat{y}_i}{\sqrt{\mathrm{var}(y_i - \hat{y}_i)}}, \tag{4.4}$$

where y and $\hat{y}$ are the predicted and test values respectively (bold case letters indicate a vector and plain letters indicate a scalar), the denominator is the standard deviation of the raw errors from Equation (4.3) (the numerator). Like in the time series models of Chapter 3, linear models assume that the variance of the errors is constant and the errors are uncorrelated. So, we are looking for whether the errors are homoscedastic, uncorrelated, and unrelated to all the covariates and fitted values[14]. In practice, some amount of heteroscedasticity always exists. For example, a situation where the homoscedasticity assumption could be violated is the volatility of annual heating load with changing envelope conductance (U-value) levels. It is to be expected that the variation of load in highly insulated buildings (due to weather) would be less than in those with low insulation levels. So, the response, and therefore the errors, could undergo a sudden shift at some critical level of insulation (U-value).

[14] For NLM, the errors should be approximately normal as well. See Davison (2003, chpt. 8) and Davison and Tsai (1992) for more details.

4.3 Step One: Classical Emulators

first... all models are wrong,
some, though, are better than others...;
and, second...
[do not] fall in love with one model,
to the exclusion of alternatives.

McCullagh and Nelder (1983), *Generalized Linear Models*

Linear models are simple, well-understood, beautiful models. In a situation where the 'data' are simulated, and where achieving an exact fit is not as important as understanding what drives the system (sensitivity, uncertainty), linear models are strong contenders. Compared to Gaussian Process regression, assessing trends comparing the relative impacts or sensitivity to different inputs is easier with classical emulators/models (both linear and non-linear). We tested three variants of linear models, discussed in this section. A short introduction to the model form is followed by the 'typical' and 'best' models for the simplest model type (Normal). Recall (from Section 4.2.5) that the 'typical' model is trained from the typical year files only, whereas the 'best' model is trained from a larger number of typical, recorded, and synthetic files. Only the NLM converged to a result with just the typical weather data for Case 1. For GLMM, only the 'best' model fits are presented.

4.3.1 Classical Linear Models

4.3.1.1 Normal Linear Models

Originally proposed by Gauss and Legendre in the early-1800s (ibidem), the simplest regression model is a linear model with Normal errors

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \tag{4.5}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T$ is an $n \times 1$ vector of responses, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a $p \times 1$ matrix of unknown parameters, $\boldsymbol{X} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{j,1} & \cdots & x_{n,p} \end{pmatrix}$ is an $n \times p$ matrix of coefficients, and ε is the $n \times 1$ vector of Gaussian errors with constant variance and zero mean.

4.3.1.2 Generalised Linear Models and Linear Mixed Models

Linear Mixed-Effects Models (LMMs) are extensions of NLMs to include both fixed and random effects. The normal linear mixed model is

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon}, \tag{4.6}$$

where $\boldsymbol{y}$ is the $n \times 1$ response vector, n is the number of observations or data points, $\boldsymbol{X}$ is the $n \times p$ design matrix of (fixed-effects) variables, $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed-effects parameters, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ error term. The new terms $\boldsymbol{Z}$ and $\boldsymbol{\gamma}$ represent the $n \times q$ random-effects design matrix and $q \times 1$ random-effects vector respectively. In this thesis we tested LMM by including two random factors (separately and together): the building type, e.g., office, apartments, and the climate/location, e.g., Geneva (This only applies to Case 2, see Table B.3.). The first is supposed to represent those aspects of a building (like compactness, exposure to wind, etc.) that are not captured by the other variables already included in the NLM (e.g., building envelope properties). The building type may be considered a random factor in a study where a stock of buildings is being modelled and the exact building type is speculative, like Case 2 here. The types used in this study are one of a very large number of possible building types (depending on the level of detail the user would like). Likewise, the climate/location is meant to represent those aspects of a climate, like proximity to water, the influence of topography, etc., that are not captured by the three variables we considered (temperature, humidity, solar radiation). Several climates could have similar values of, e.g., mean annual TDB, but not similar seasonal characteristics (a climate with hot summer and cold winter would have a similar annual mean to a climate with mild seasons). We will see later that this did not confer a significant advantage.

The Generalised Linear Model (GLM) is an extension to NLM that allows the relaxation of several assumptions, chiefly that the probability density of the response does not have to be normal, and the mean response may be related to the linear predictor ($\boldsymbol{X\beta}$) by a monotonic function. The results from fitting GLMs (or GLMM) are not presented, since they do not provide an improvement over NLMs (or LMM). Candidate output distributions such as the Gamma and Inverse Gaussian were considered and rejected. We did not find a reason to justify using a different distribution for the output, either in the literature or in our own analysis.

4.3.2 Analysis of Variance (ANOVA)

A simple Analysis of Variance (ANOVA) with Gaussian errors is way of representing outputs from a classical linear model (Section 4.3.1.1). An ANOVA with *fixed* and *random* effects is the same for the LMM (Section 4.3.1.2). We initially examined the different inputs by using n-way ANOVA[15]. ANOVA is a robust tool for exploratory analysis, though it is not expected to do well for highly non-linear systems, or when the outputs are non-normal. However, it can offer useful pointers on which factors bear examination. Given that this preliminary ANOVA-based analysis did not play a significant role in the development of the regression models, we do not present results.

In the analysis, all factors, except those representing building envelope properties, were treated as being continuous. The envelope factors were discretised into bins for analysis to avoid making a distinction between, for example, $U = 3.25W/m^2K$ and $U = 3.26W/m^2K$. The envelope factors were tested under two conditions: once treated as 'random' and otherwise as 'fixed'. Depending on the kind of analysis being undertaken, they can be treated as both. In a design exercise, the envelope properties are fixed because they represent specific values that the user would like to test. If the user is modelling existing buildings for forensic analysis, then the envelope properties observed (from experiments or survey) are random. Being treated as random factors put a stronger condition on factors to be significant, which (in this example) meant they usually were not significant. A test in which these factors are treated as fixed effects, that is to say they are not random samples of an infinite population but actual fixed levels of interest, changes the p-value to almost zero (i.e., high significance in determining the variance of the output).

4.3.3 Results and Discussion: Linear Models

4.3.3.1 Guide to Linear Regression Plots

Two types of plots are used to present the regression results: 'prediction' plots (e.g., Figure 4.7a), and 'transversal' error plots. The transversal plots show the evolution of error with increasing training samples (e.g., Figure 4.26). The graphs have all been calculated using the data presented in Figure 4.4, and may be reproduced using the scripts posted with the archive copy of this thesis (infoscience.epfl.ch).

In the prediction plots, 'predicted' values are plotted against 'known' or 'simulated'

[15] MATLAB function `anovan`.

values. Known/simulated values are plotted on the x-axes for all regression figures. These values are ‘known’ in the sense that they were the obtained from a full Energy Plus simulation. Values ‘predicted’ by the regression model are plotted on the y-axis. Results from the training and testing data sets are plotted side-by-side, with heating above and cooling below. Plotting the training data is a self-validation to see the goodness-of-fit, while the plot against testing data is a test on nominally ‘unseen’ data. Additional plots examine the errors from the regression fits for Normality and homoscedasticity (e.g., Figure 4.8).

Generally, the plots are presented in terms of the z-scores. The reverse transformation, i.e., recalculation from z-scores back to the original quantities, created no perceptible distortions for any of the models. Consequently the results are presented solely in terms of the z-scores. This section presents NLM fits for Case 1 (Geneva home) and Case 2 (USDOE buildings), and a LMM to Case 2 only. All fit types (typical, second, and best) are included, except where a fit could not be calculated because of insufficient variation in the training data (e.g., LMM for Case 1).

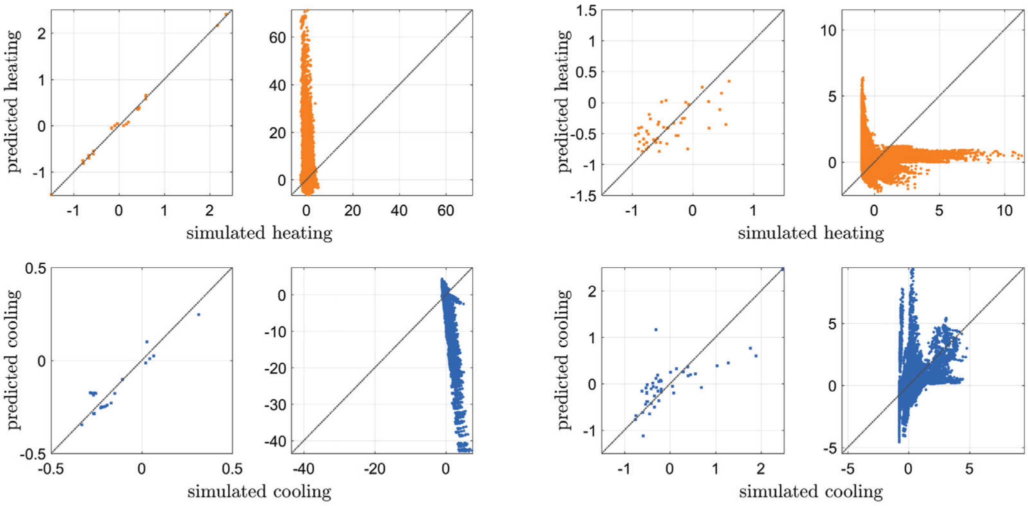
4.3.3.2 Prediction and Error Plots

Examining the plots presented in Section 4.3.1, the performance of the linear models is less robust than expected, given their ubiquity in building simulation literature (Section 2.5). The model based on typical data only is clearly unable to explain the data obtained from other weather files (Figure 4.5) – at the minimum at least additional simulations with recorded files are needed. One can see from the right-hand side sub-plots of Figure 4.5, where the model is attempting to predict unseen (testing) data, that the predictions are nearly flat lines. The prediction intervals were so wide as to be useless, so are not plotted here. Results with more training data are better (second and best fits), and this behaviour is common to the other model types. For the restricted sensitivity analysis, Case 1, the NLM performs acceptably with random data (Figure 4.7), achieving its best fit with about 200 training data points (Figure 4.25a). Both the typical and second models considerably over-fit to the training data, resulting in wayward performance and errors (Figures 4.5 and 4.6). The picture is similar with NLMs for Case 2 (Figure 4.7b). Introducing a random factor (building type) for Case 2, i.e., a LMM model, does not improve the fit significantly either (Figure 4.10). Some additional plots examine the properties of errors (Figures 4.8 and 4.9). We are looking for the errors to be nearly Normal (with a Quantile-Quantile plot, or *qqplot*) and homoscedastic, i.e., that they have nearly equal variation when plotted against the corresponding prediction. The latter is to check that the errors are not somehow being affected by, or showing correlation with, the predictions.

The transversal error plots in Figure 4.26 show that, regardless of the quality of the fit achieved, the error tends to plateau after the training data set exceeds approximately 200 points. The errors for the NLMs stabilise with 200 or 500 training points (cases 1 and 2 respectively), which are very small compared to the testing set (approx. 0.5 % and 0.6% respectively). The GLM and GLMM errors show similar behaviour for Case 2, though the GLM for Case 1 does not converge until about 600 training points. Case 1 always needs less training data than Case 2, since it is trying to model a far more limited situation (one building – one climate).

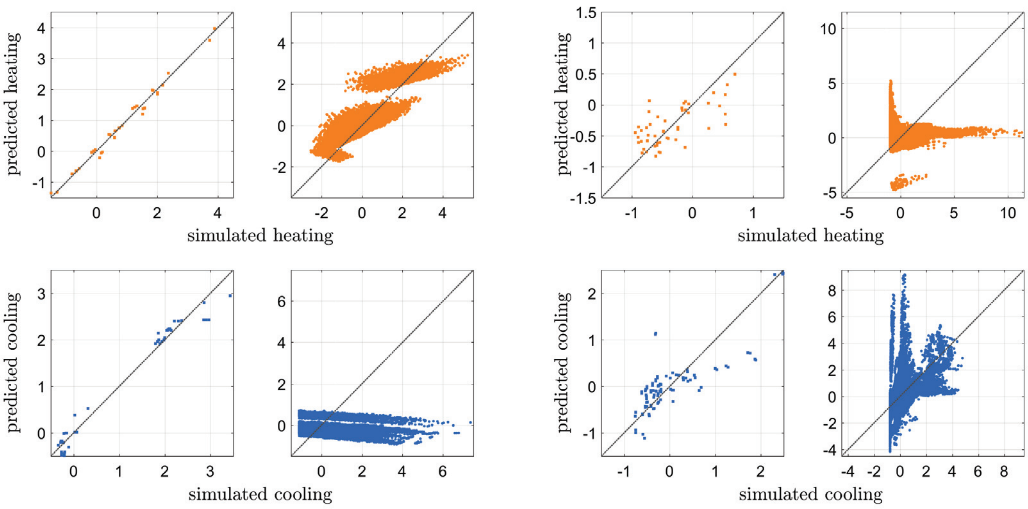
As we have argued before, the underlying data generating process that is being modelled – building simulation, or measured building data[16] – has a non-linear response to changes in most independent variables (design inputs). Linear models seem to work acceptably for restricted sensitivity analyses of simple buildings (e.g., a single-family home) to climate variation in a limited range (Figure 4.7a). It is not particularly surprising that the random sampling iterations find the 'best' model (even if the best model is not particularly good at prediction) with a relatively small number of training data points, compared to the size of the unseen/testing data. This is because we are dealing with annual aggregate descriptors of the climate like median (Table 4.1), so the complete range of possibilities is sampled relatively early on. If one were dealing with an emulator of building simulation or operation trying to predict at the hourly time scale, like those proposed in Sections 5.3.4 and 5.3.7, the number of training data points would presumably be much higher because of the far higher number of possible combinations of temperature and other parameters over time.

[16]Since, presumably, the ultimate goal is to model the functioning of *actual* buildings, random occupancy, construction errors, atypical weather, etc.



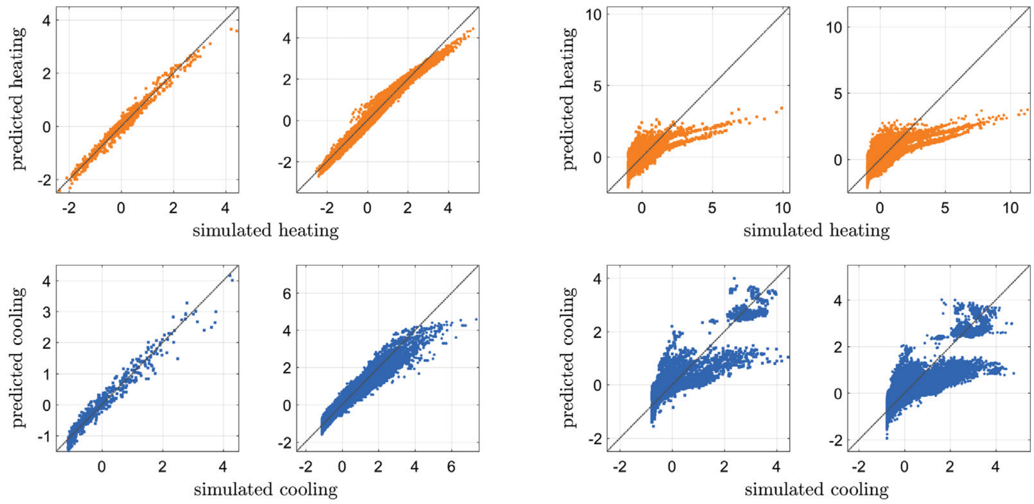
(a) *Case 1 – Geneva home* **(b)** *Case 2 – USDOE buildings*

Figure 4.5 – *Typical fits, Normal Linear Model (NLM), Cases 1 & 2. Clockwise from top-left for each case: training data for heating, testing data for heating, testing data for cooling, training data for cooling. These are z-scores, not the original data.*



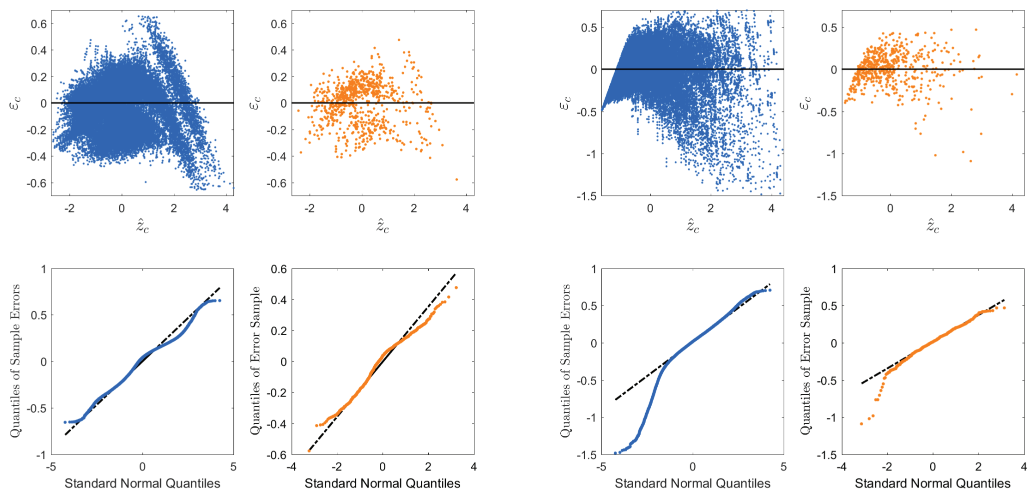
(a) *Case 1 – Geneva home* **(b)** *Case 2 – (USDOE) buildings*

Figure 4.6 – *Second Fits (typical and recorded data), NLM, Cases 1 & 2. Training data is plotted on the left, testing on the right, heating on top, and cooling on the bottom. These are (unit-less) z-scores, not the original data.*



(a) *Case 1 – Geneva home* **(b)** *Case 2 – (USDOE) buildings*

Figure 4.7 – *Best Fits (random training data selection), NLM. Clockwise from top-left: training data for heating, testing data for heating, testing data for cooling, training data for cooling. These are z-scores, not the original data.*



(a) **(b)**

Figure 4.8 – *Prediction errors from the Best Fit, Case 1, NLM – heating [left] and cooling [right]. The subscript 'c' indicates cooling, and 'h' is for heating. The ε is calculated as shown in Equation* (4.3)*. The errors are neither Normal (see the qqplot), nor homoscedastic (the errors 'fan out' for higher values of predictions). These are z-scores, not the original data.*

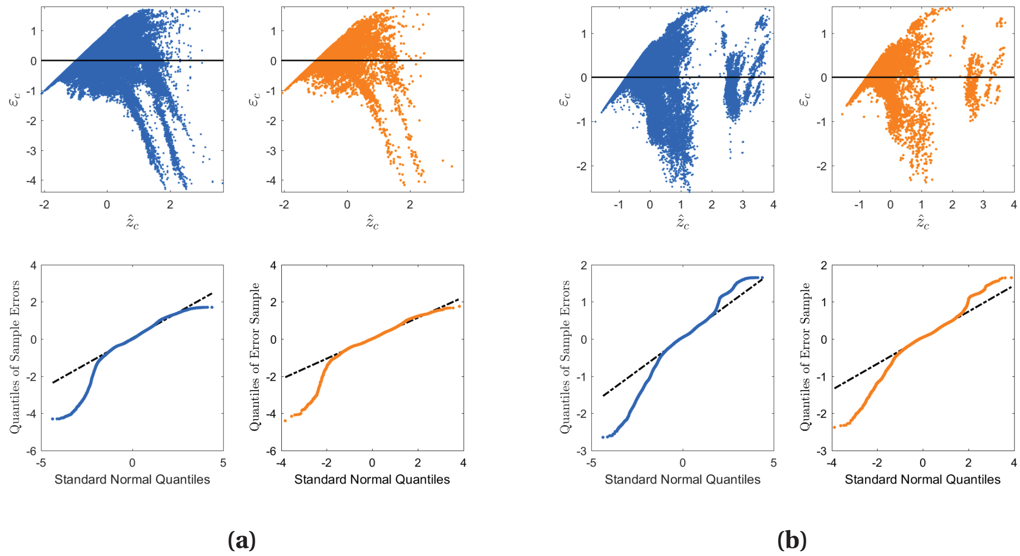


Figure 4.9 – *Errors from the Best Fit, Case 2, NLM – heating [left] and cooling [right]. The errors are neither Normal (see the curve in the qqplot), nor homoscedastic (the errors 'fan out' for higher values of predictions). These are z-scores, not the original data.*

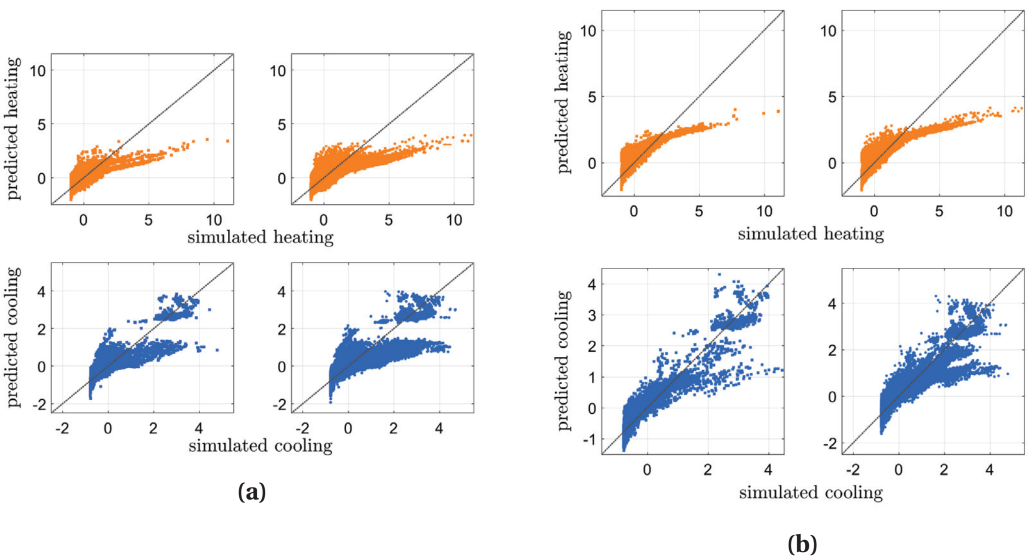


Figure 4.10 – *Predictions from LMM model for Case 2 only – Second Fit [left], Best Fit [right]. Setting the building type to be a random factor, in addition to the same fixed factors used in the NLM before, did not improve the fit. Training data on the left and testing on the right of each sub-plot. These are z-scores, not the original data.*

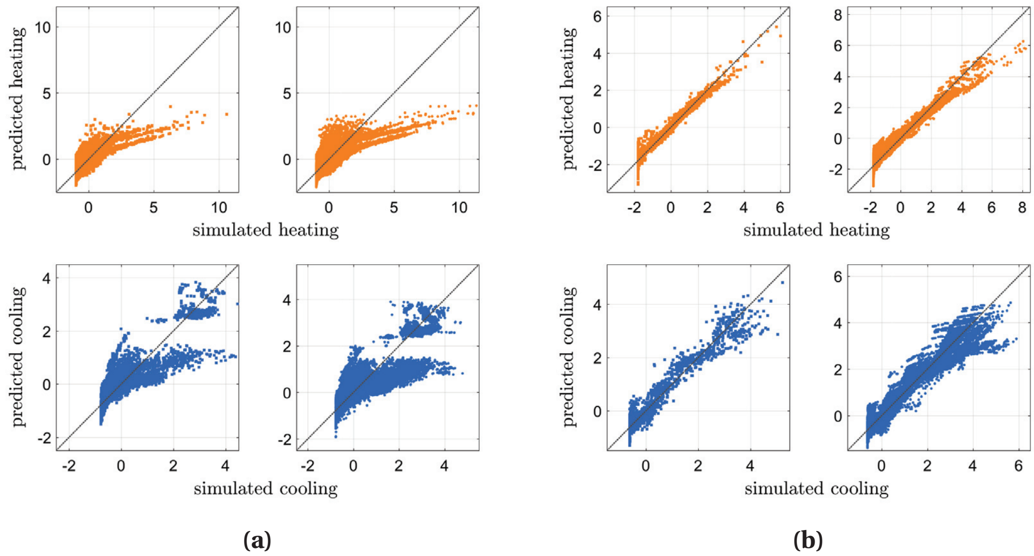


Figure 4.11 – *Predictions from NLM model for subsets of Cases 1 & 2 only, Best Fit. The left sub-plot is for the 'office' building type from the Case 2 dataset (see Table B.3). The right sub-plot is for the single-family home (Case 1) in all climates (Section B.4). These are (unit-less) z-scores, not the original data.*

4.3.4 What about Non-Linear Fits?

We attempted to fit non-linear modifications of the input parameters like higher-degree polynomials and exponentials. Given the complex shape of the underlying data along individual dimensions in a multi-dimensional space, it is very difficult to select the functions. It is possible, with many software packages, to apply a brute force method where a very large set of possible linear and non-linear terms are combined to see which blend works on the given data set. We do not prefer this method for several reasons: it is difficult to generalise, since the form of the function will inevitably vary based on the training data set; and because non-linear models in multiple dimensions tend to obscure the effect of individual factors. In both the time series models of Chapter 3 and the regression models here, we are aiming for a stable form of the model, where the coefficients of the model parameters may change with each problem but not the structure. Naturally, the nature of the problem dictates which parameters stay in or are thrown out, but in a non-linear model, they may be used in a polynomial or exponential of any degree, thereby changing their influence on the outcome substantially. The preferred solution, Gaussian Process regression, is not as simple to interpret as a linear model, but the hyper-parameters (length scale and noise standard deviation) and the form of the kernel function *are* indicative of the nature of the output data.

4.4 Step Two: Probabilistic Emulators

After discussing the performance of classical parametric methods (linear models) in Section 4.3.1 and considering (but discarding) classical non-parametric approaches, we proceed to test a relatively novel non-parametric regression technique: Gaussian Process regression[17]. The advantages of this technique for our application include: relatively small training data sets for equivalent performance, explicit statement of prediction intervals, and an inclusion of the influence of each variable in the model (through the length scales discussed below). See Rasmussen and Williams (2006, chpt. 1) for a discussion of why probabilistic fitting approaches, like Gaussian Process regression, are better than those based on a restricted class of functions like linear models.

[17] Novel for building simulation, that is (see Section 2.5).

4.4.1 Gaussian Processes and Regression

4.4.1.1 A Very Brief Introduction

Gaussian Process regression is the indirect search for an underlying *latent function* to describe non-linear data Rasmussen and Williams (2006) describe it as supervised learning of "input-output mappings from empirical [continuous] data (the training dataset)"[18]. It involves the fitting of two functions to the data: a mean function [$\mu(x)$], assumed to be zero in this application, and a covariance function [$k(x_i, x_j)$]. This (covariance) function does not, by itself, predict the outcome at unseen points. Instead, it defines the change of the parameters of a GP over a basis space defined by the inputs. This implies that, at any new/unseen query point, the Gaussian Process regression fit will give the parameters of a univariate Gaussian distribution (mean and variance), where these parameters are related by a covariance function. The mean of the univariate Gaussian distribution at a particular query point is not the same as the mean function describing the entire data. Rather, it is the mean of the output, a random variable, at a given combination of inputs. Hence, the regression prediction or output at any given combination of independent variables is a Gaussian random variable.

"A Gaussian Process (GP) is a generalisation of the Gaussian probability distribution" (ibidem)[19]. The difference is that while a probability *distribution* describes "random variables which are scalars or vectors", a probabilistic *process* "governs the properties of functions" (ibidem). In practice, we do not work with functions, treating them instead as "very long vectors" (ibidem). This makes the process far simpler to implement in finite-precision arithmetic because we can sample the functions discretely and still get their correct properties. In other words, at every point, or finite collection of points, the function responds as if the entire (infinite) population of points described by that function had been considered (ibidem). And these properties are consistent with queries from any other sample. So, every data set $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ may be thought of as a "single point sampled from some multivariate [d-dimensional] Gaussian distribution" (Ebden 2008). A GP is completely specified by its mean function

$$E(f(x)) = \mu(x), \tag{4.7}$$

[18]The same process for discrete outputs would be called *classification*.

[19]Alternatively, a "Gaussian process generates data located throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution" (Ebden 2008; The MathWorks, Inc. 2015).

and its covariance function

$$\mathrm{Cov}(f(x_i), f(x_j)) = k(x_i, x_j), \tag{4.8}$$

where $f(x)$ is called the *latent function*. The covariance function is usually modelled as a function of some parameter vector $\boldsymbol{\theta}$, known as the vector of *hyper-parameters*. Practically speaking, it is very useful to assume that the underlying GP has zero mean everywhere[20]. Once the mean is accounted for, the observations are related only by a covariance function $k(x_i, x_j)$. The underlying structure of the data should be reflected in the choice of a covariance function or *kernel*: is it smooth, periodic, linear, some combination of these, etc.? If classical regression is the direct search for some function $f(x)$[21], Gaussian Process regression is the search for this covariance function.

Gaussian Process regression is almost identical to Kriging, and some foundational formulations and assumptions of both are presented in Section B.2. There are some differences between how Kriging or Gaussian Process regression is used in practice, so only the latter term in used in this thesis. The references we used[22] tend to use varying terminology and interpretations to introduce the concept of GPs, and Gaussian Process regression in particular. We will try to be self-consistent, which means we will not manage to follow any of the sources consistently. Only some of the fitting results and a discussion of the model performance are presented in this chapter. For additional issues and discussion, see Section B.2.

4.4.1.2 K-fold Cross-Validation

K-fold cross-validation is a simple intermediate step to train robust regression models, by looking for the best hyper-parameters (length scale and noise standard deviation) for a GP fit[23]. Essentially, k-fold cross validation yields a pre-selection of the best hyper-parameters, based on subsets of training data, that are expected to deliver the best performance over the whole training data set. Recall that the simulated data set was already split into training and testing sets to test the predictive accuracy of the model at the outset. K-fold cross-validation is only implemented on this ‘master’

[20] Analogous to *simple kriging*.

[21] Which is thought to underlie the observations $\boldsymbol{y}$, such that $y = f(x) + \varepsilon$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

[22] Christensen (1991), Duvenaud (2014), Ebden (2008), Kleijnen (2009), Mackay (1998), Rasmussen and Williams (2006) and The MathWorks, Inc. (2015)

[23] Hyper-parameters describe the covariance function, not the function modelling the output directly, so this is not the same as the step where we find model coefficients in classical regression or time series models (Section 3.5.2).

training set.

In k-fold cross validation, the training data is further split into k subsets. Then, a specific model (i.e., a model based on specific hyper-parameters) is fitted to $k-1$ subsets (mixed together) and tested for its predictive accuracy on the k^{th} subset (using RMSE, for example). In our implementation, we use two nested k-fold loops, with $k = 10$ (for both outer and inner loops). The number of folds is an arbitrary choice, so the calculations would finish in time. Future work could optimise this based on computing capacity. The outer loop is to select the 'best' model, i.e., one with the minimum RMSE. The inner loop is to estimate the average RMSE over different combinations of training and testing k subsets, to enable this comparison of RMSE values.

In each iteration of the inner loop, the RMSE is calculated and stored separately for each k^{th} subset, using the same hyper-parameters (we are testing only one combination of hyper-parameters over one iteration of the outer loop). This gives an average RMSE (calculated over the iterations of the inner loop) for each iteration of the outer loop, i.e., k different outer-loop averages for k different hyper-parameter combinations. These inner-loop averages are compared at the conclusion of the outer loop, and the hyper-parameters with the minimum average RMSE are selected to fit the overall model. For each iteration of the outer loop, the hyper-parameter for length-scale is one element of a k-dimensional vector of fractions of the mean Euclidean distance between all input points[24]. The second hyper-parameter is also one element of a k-dimensional set of fractions, this time of the standard deviation of the training output data points (regressands). The procedure is summarised below.

1. Working with the master training data, calculate the mean Euclidean distance between all inputs and the standard deviation of outputs.
2. Split the data into k subsets. Select hyper-parameters to test on this subset. Length scale comes from the mean Euclidean distance of inputs, and the variance parameter is a fraction of the standard deviation of the outputs.
3. Fixing the hyper-parameters, fit a model to each group of $k-1$ subsets and test on the k^{th} subset (i.e., leave out a different subset each time, repeating for all subsets to get $k = 10$ different fits). Calculate the averageRMSE for each fit.
4. Compare the average RMSE from each model, i.e., hyper-parameter choices. Select the hyper-parameters that give the lowest average RMSE.

[24] For the automatic relevance determination (ARD) kernels, multiple random selections are made from this set of fractions.

4.4.1.3 Choice of Kernels

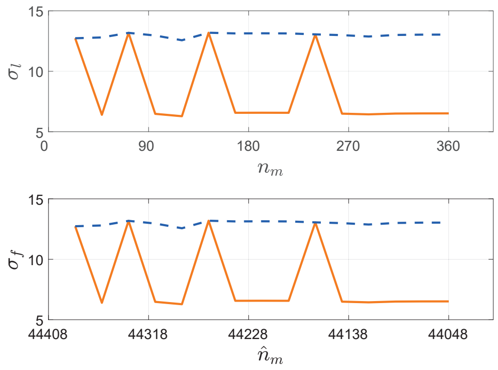
We tested the performance of two kernel functions, from the very large variety available (see Section B.2.1). The choice of kernels in Gaussian Process regression is like the choice of model structure in classical regression. Except, instead of trying to estimate the shape of the actual response variables, we are now trying to *model the form of the variance due to each input and their covariance.* The covariance function specifies the covariance between the values of the latent function at two different points, say $f(x_i)$ and $f(x_j)$. The kernel is also an estimate of roughness, like a classical regression function, but of the covariance instead of the independent variables. It can only be a positive semi-definite function[25]. The literature, as we understand it, never explicitly recommends a 'default' or 'best' kernel. This would be as meaningless as specifying a default linear model for any data set in any problem. The kernel has to reflect the underlying distribution and one's knowledge of it. Hence, it is the toughest decision to make in the whole process. The examples of kernels shown here are limited, and *as simple as possible*, prioritising simplicity (under-fitting) over over-fitting with a complicated kernel.

The squared exponential function (SqE) is probably the default choice for most applications, because it works reasonably well for "interpolating smooth functions" (Duvenaud 2014). It assumes that the underlying function ($f(x)$) of the response variable is infinitely differentiable, and that there are no "kinks" in the function. If the function or one of its derivatives has a discontinuity, or even sharp roughness, then the length scale ends up being determined by the shortest "wiggles" in the function. That is to say, the GP tries to fit to the finest features, and ends up performing badly on the broadly smooth parts. In multidimensional data, it is not obvious whether such features exist, so we applied a test advocated by Duvenaud (ibidem): checking the evolution of the length scale (Figure 4.12).

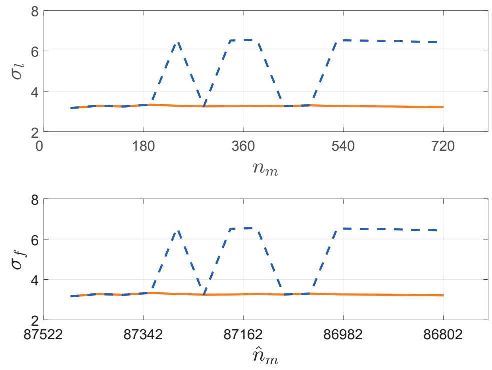
Duvenaud (ibidem) recommend that if the length scale keeps getting smaller with more data, then such roughness features likely exist and the kernel is inappropriate. This is unclear in Figure 4.12, where the length scales are neither increasing nor decreasing consistently. The cooling model is less stable than the heating model for Case 1, but more stable for Case 2. Based on this, and the fact that we expect the different types of inputs to cause different types of non-smoothness in the output, we tested the ARD kernel[26], which over-fits to the training data and predicts badly on unseen data (Section B.2.4). We did not attempt to use other kernels in this thesis,

[25] See Rasmussen and Williams (2006, chpt. 4) for an explanation of what that means.
[26] See Section B.2.1 for how the ARD kernel is different from the SqE.

though that would be an avenue for future exploration. The work of Duvenaud (2014) and Duvenaud, Lloyd et al. (2013) could be a useful basis for an exploratory kernel-selection step in developing Gaussian Process regression models. Fits using the SqE kernel are presented in this chapter and in Section B.2.3, while fits using the ARD kernel are presented in Section B.2.4 only.



(a) *Case 1 - Geneva home.*



(b) *Case 2 - USDOE buildings.*

Figure 4.12 – *The length scale, σ_l [top], and variance, σ_f [bottom], plotted against sizes of training (n_m) and testing ($\hat{n}_m$) data sets, respectively. The dashed blue line is for cooling models, and the solid orange line is for heating. The length scale is not consistently decreasing with more data, especially in the first case. Results are from using the SqE kernel.*

4.4.2 Results and Discussions: Gaussian Process Models

4.4.2.1 Guide to GP Plots

Two kinds of plots are presented for the Gaussian Process regression examples: prediction-error plots (e.g., Figures 4.13, 4.14, 4.17 and 4.18) and prediction interval-relative error plots (e.g., Figures 4.15, 4.16, 4.19 and 4.20). The prediction from Gaussian Process regression is the same as that from linear models, and consists of mean response at a specific combination of independent variables. The error is once again the difference between the regression prediction and 'true' value for the testing set, like in Equation (4.3). In the prediction interval-relative error plots, the intervals plotted around each prediction value are prediction intervals on the regression prediction, since the prediction is a Gaussian random variable. The relative error is the ratio of the error at a specific combination of independent variables to the predicted mean value at that point. Note that the z-scores are unitless, and ε always represents a prediction error.

First, examine the prediction-error plots. The left-hand sub-plots of this group show simulated against predicted values[27]. The 1:1 line is plotted as a guide. The number in the top-left indicates the number of training data points. The right-hand sub-plots show the distribution of errors. In the prediction interval-error plots, the lines in the left-hand series of sub-plots represent a prediction interval ($\pm\sigma$, where σ is the variance given by the Gaussian Process regression fit). The simulated values are represented by black dots. Only a hundred predicted-simulated pairs are plotted to avoid overcrowding, along with their prediction intervals. In most plots, it looks like too few black dots are outside the prediction interval (roughly 32% should be out and 68% in, since we are plotting the extent of one standard deviation). Only 100 testing points were randomly selected to be plotted, whereas the 68% figure arises when the whole sample of testing points are considered. The right-hand sub-plots show histograms of the ratio of each error to its corresponding prediction, the relative error. The relative error was censored, since the ratio of one error-prediction combination where the denominator is small enough would produce a very large number with a very small probability, making a visual appraisal of the histogram very difficult. The cut-off values were the 97.5$^{\text{th}}$ and 2.5$^{\text{th}}$ percentiles. Note that the frequencies (y-axis) of histograms are always stated as probability densities.

4.4.2.2 Plots: Squared Exponential Kernel

Like for the linear models(Section 4.3.3), the Gaussian Process regression models trained on randomly selected data set do better than those trained on typical or typical and recorded weather files. This is seen in the first iteration already (with only 24 randomly selected files), improving slowly thereafter. Looking at the plots presented in Section 4.5, one case see that the quality of prediction that was achieved by the linear models with about 200 training points is achieved with less than 48-120 points for heating and cooling in Case 1. For Case 2, the corresponding training set sizes drop from more than 500 to 200-250. As expected, typical data alone did not predict results for other weather conditions satisfactorily, and a random selection of weather files (inputs) for training performs best. This result reinforces the finding, first presented in Section 4.3.3, that there are is not enough information contained in a typical file to train the model for the complete range of weather conditions. The quality or type of regression technique is irrelevant if the training data is not representative enough of the testing data. The same number of random training data points do better than the typical data alone for all model types. The graphs have all been calculated using the

[27] Recall that 'predicted' refers to values predicted by the regression equation, while 'simulated' refers to values from a full simulation in EnergyPlus.

data presented in Figure 4.4, and may be reproduced using the scripts posted with the archive copy of this thesis (infoscience.epfl.ch).

The results from Case 1 are better than those from Case 2, as expected. The plots of Case 2, Figures 4.17 and 4.18, show a significant amount of spread in the predictions. However, the number of off-diagonal points, i.e., predictions with high errors, is small. This is evidenced by the shapes of the error histograms with each simulated-prediction plot: the errors cluster around zero, with some outliers. A similar picture is evidenced by Figures 4.19 and 4.20, where the majority of simulated values fall within the prediction interval of the predicted values. The results from Case 1 are also easier to interpret. Note the noticeable decrease in the range spanned by the prediction interval in the left-hand sub-plots of Figures 4.15 and 4.16, for example. Essentially, the model is more 'sure' about the predictions with increasing data. The ratio of the errors to their corresponding predictions has a similar shape to the raw errors, but a small number of the ratios seem to be very high. This is a common problem when dealing with small-magnitude quantities – if the denominator is close to zero, then the fraction is going to become large, regardless of the size of the numerator.

Generally, across all cases, cooling load does worse than heating load. This is most obvious in the plots for Case 1 – Geneva home. We expect that this behaviour occurs due to two reasons: the cooling load values tend to be small, and a majority of them are zero; and the distribution of the cooling load in this climate (heating dominated with mild summers) is strongly non-normal (exponential, see Figure 4.4a). Essentially, if the model is trained on a large number of zero values, it is not going to give high non-zero values for other query points. In Case 1, some future weather files and renovation cases do produce high cooling load values, but the majority of the combinations show none. Cooling predictions are a little better for Case 2 due to the presence of hot climates (e.g., Delhi, Mumbai, and Phoenix, see Tables B.5 to B.8). Conversely, the presence of these climates, where heating loads are either zero or very small, 'misleads' the heating model (Figure 4.17). Comparing cases 1 and 2, we see that the inclusion of a multitude of buildings and climates creates a worse model. In addition, these general models require more data to train. This is why we do not suggest the use of one-size-fits-all, or unified, regression models.

Variations or subsets of Cases 1 and 2 are presented in Figures 4.11 and 4.21 to 4.24: the single-family home in Case 1 and the office building in Case 2 in *all* of the climates in the test set (Tables A.1 and B.5 to B.8). The regression plots of the modified cases show differences of predictive performance that are comparable to the original cases. That is, Gaussian Process regression predicts better, with fewer training data points.

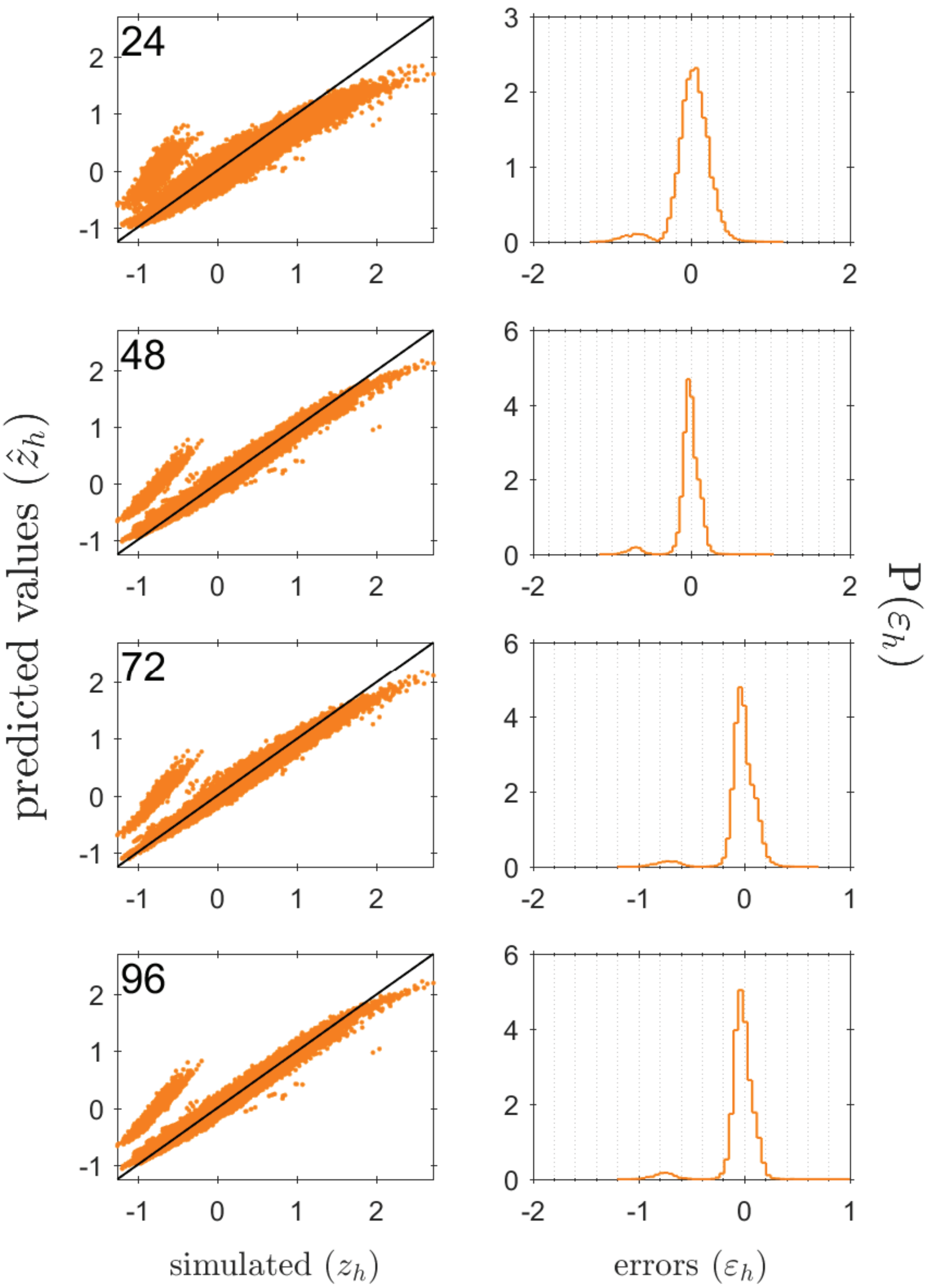


Figure 4.13 – *Case 1, Best Fit – heating. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. These are (unit-less) z-scores, not the original data.*

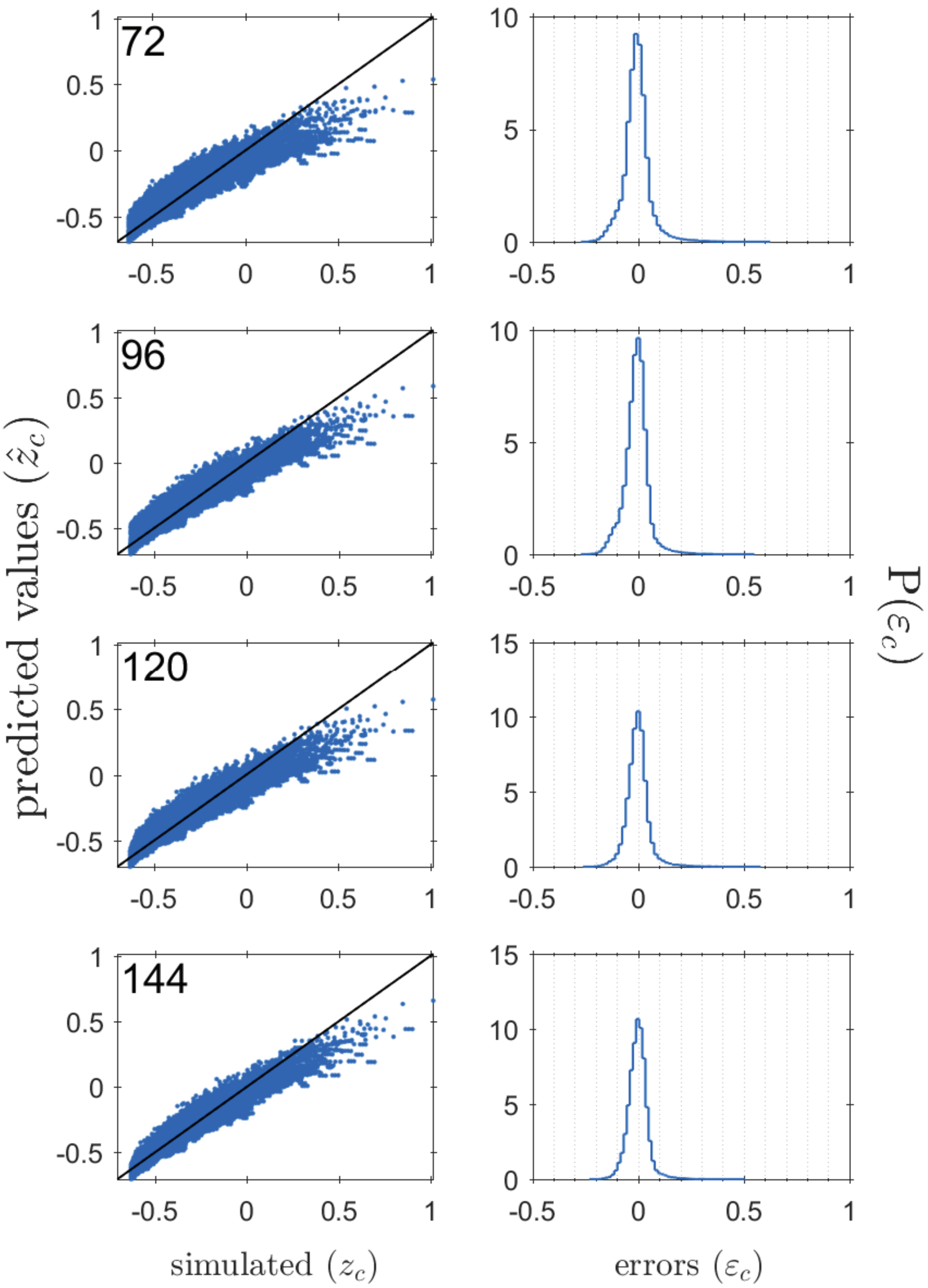


Figure 4.14 – *Case 1, Best Fit – cooling. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. These are z-scores, not the original data.*

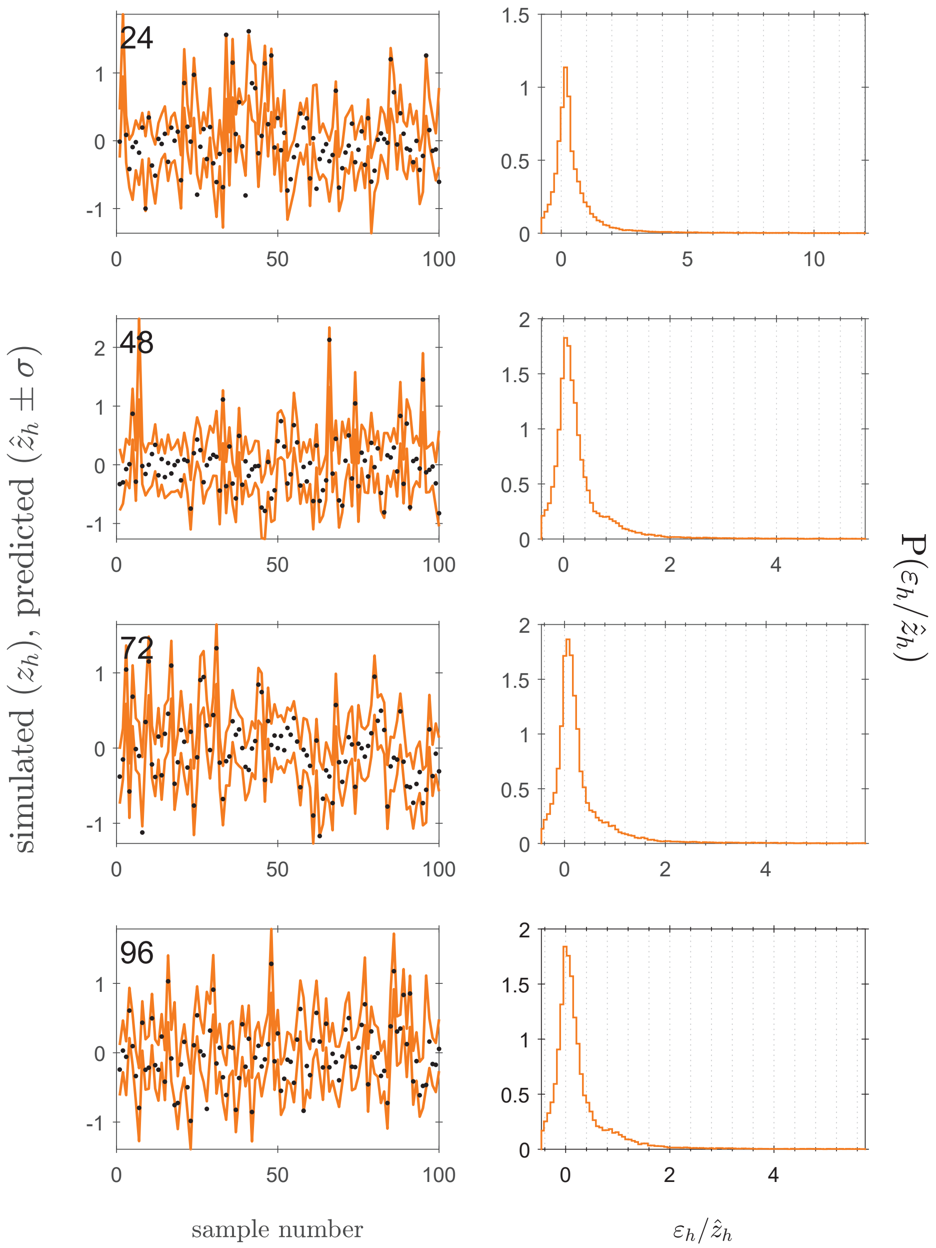


Figure 4.15 – *Case 1, Best Fit – heating. Prediction interval (68%) of predicted values enclosing simulated values (black dots) [left]. Ratio of errors to predictions [right]. These are z-scores, not the original data.*

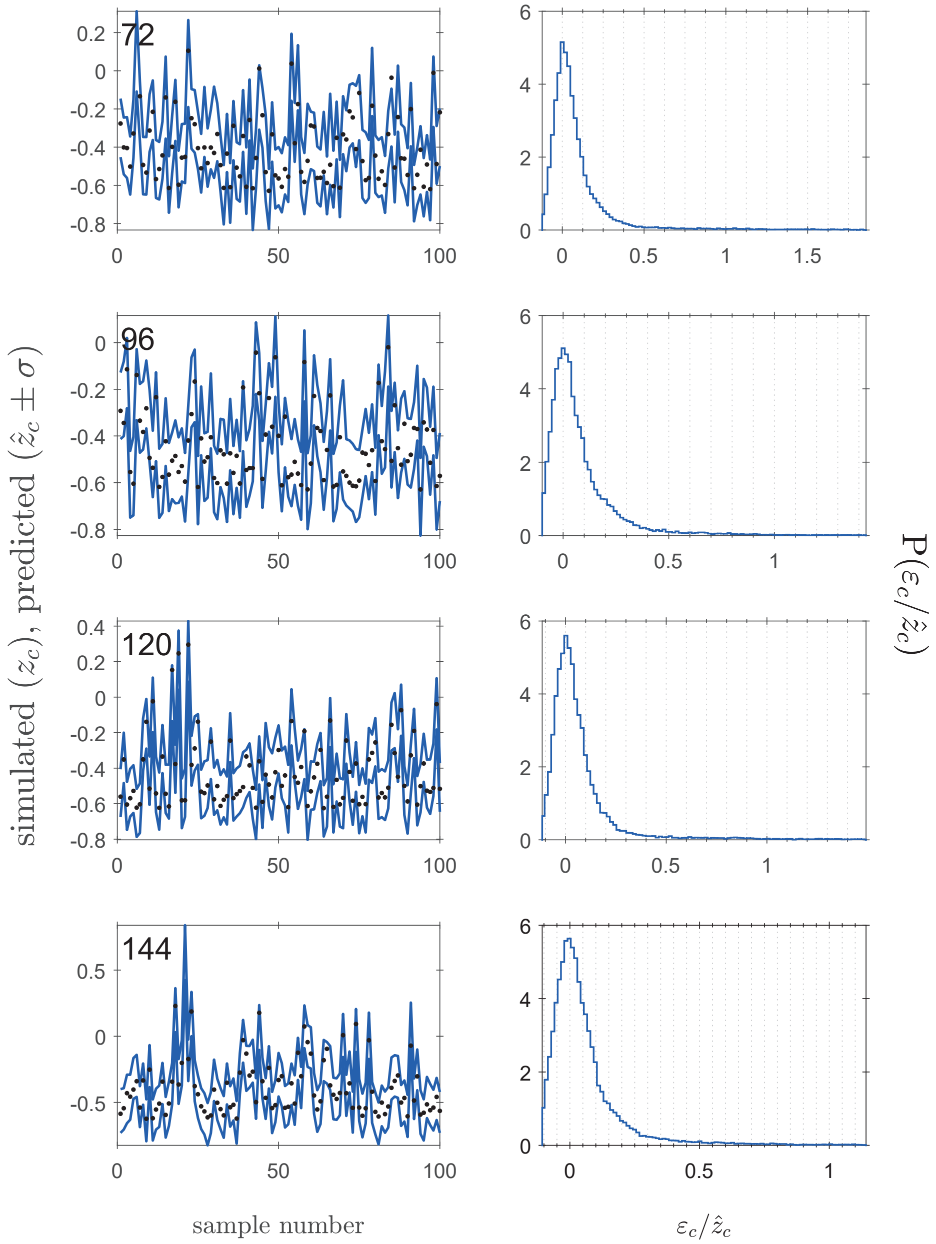


Figure 4.16 – *Case 1, Best Fit – cooling. Prediction interval (68%) of predicted values enclosing simulated values (black dots) [left]. Ratio of errors to predictions [right]. These are z-scores, not the original data.*

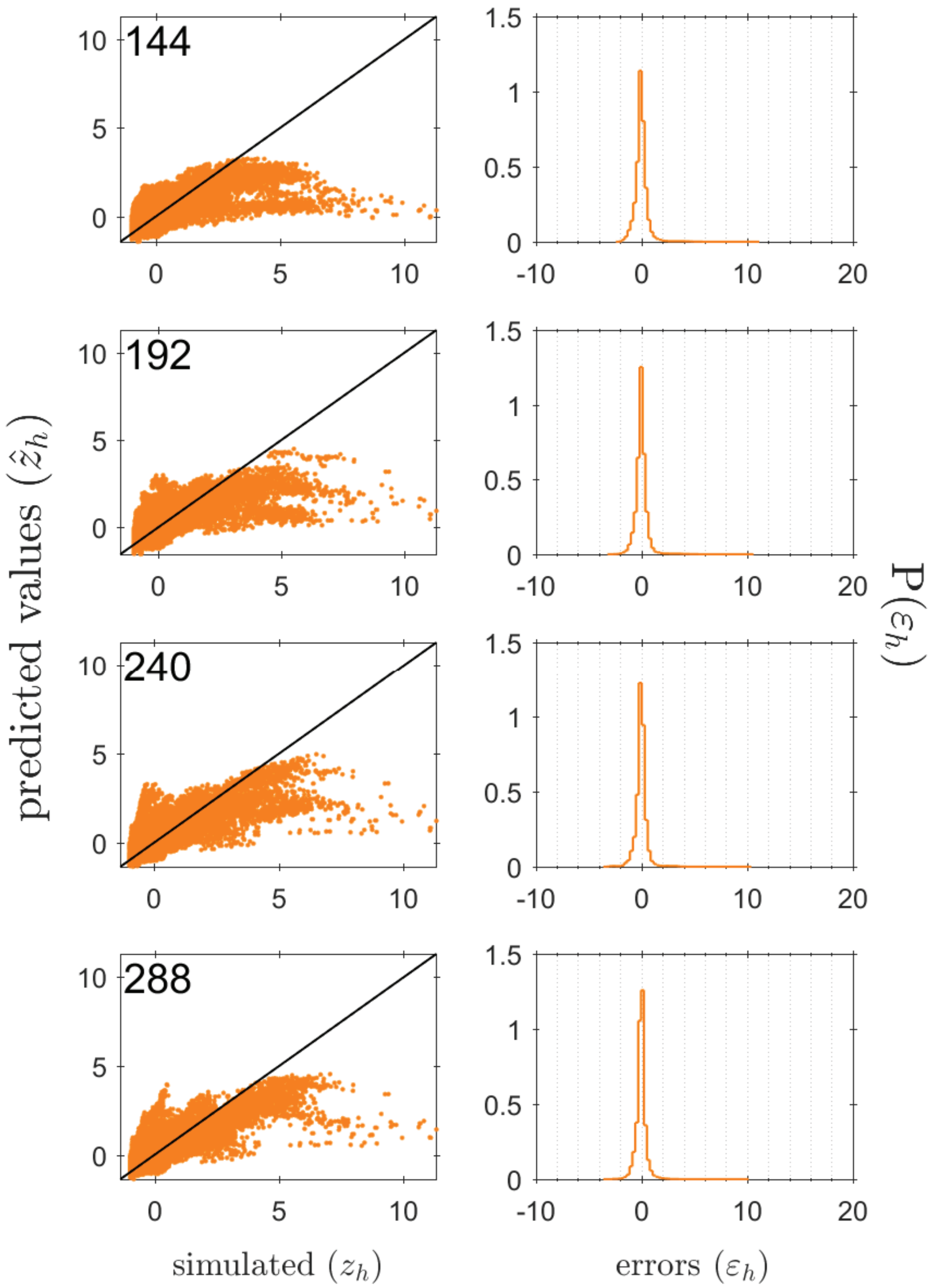


Figure 4.17 – *Case 2, Best Fit – heating. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. These are z-scores, not the original data.*

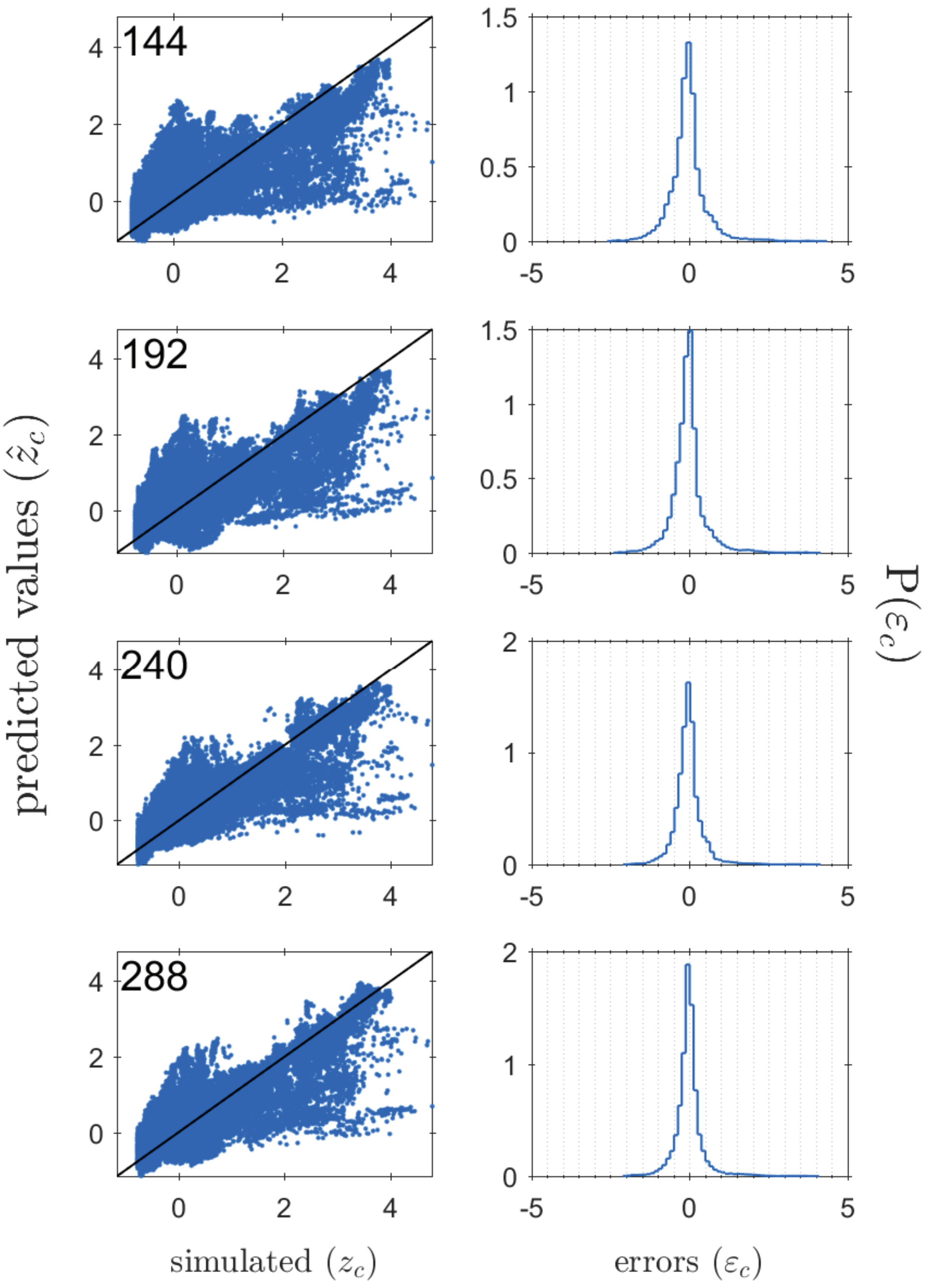


Figure 4.18 – *Case 2, Best Fit – cooling. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. These are z-scores, not the original data.*

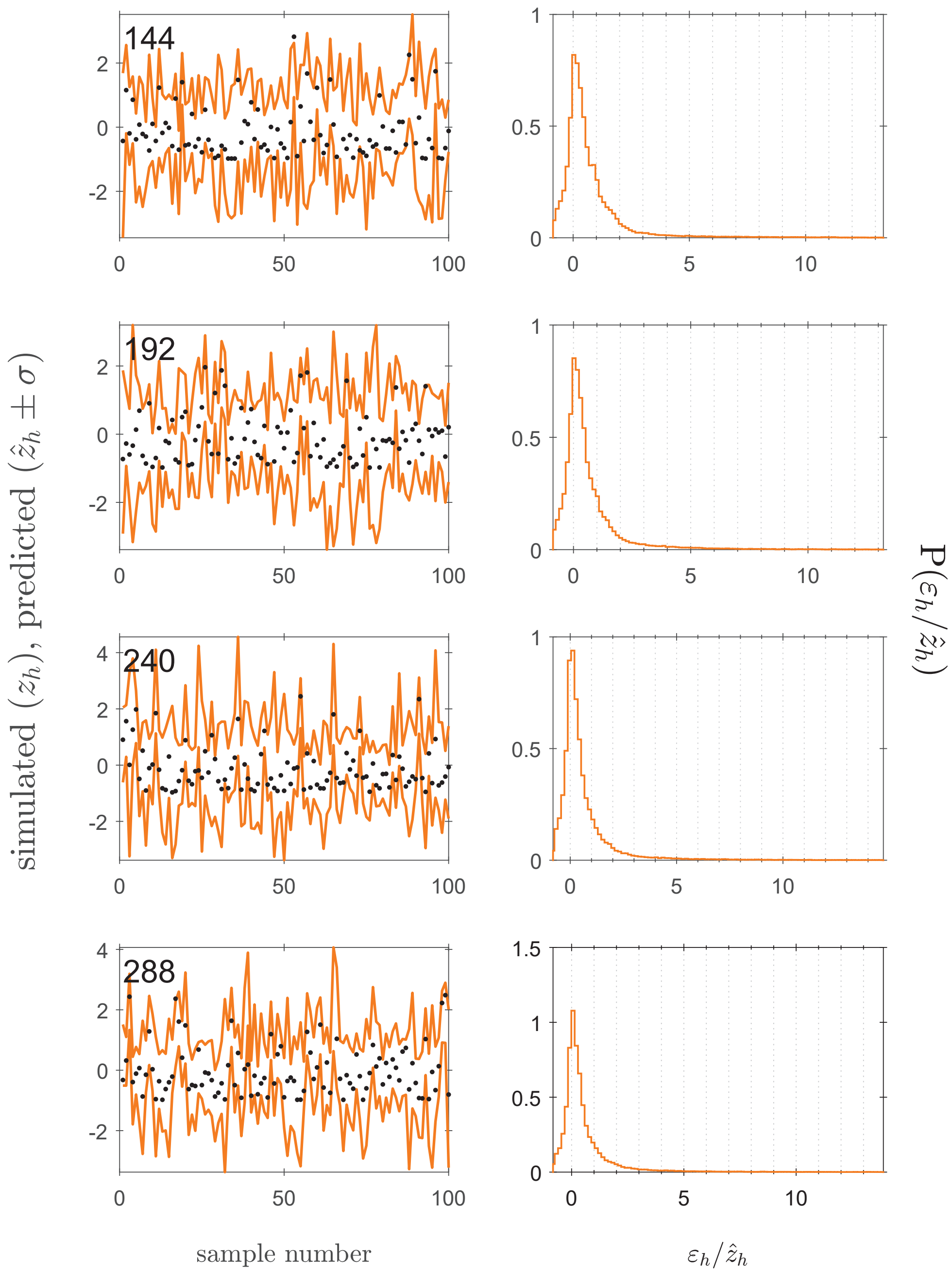


Figure 4.19 – *Case 2, Best Fit – heating. Prediction interval (68%) of predicted values enclosing simulated values (black dots) [left]. Ratio of errors to predictions [right]. These are z-scores, not the original data.*

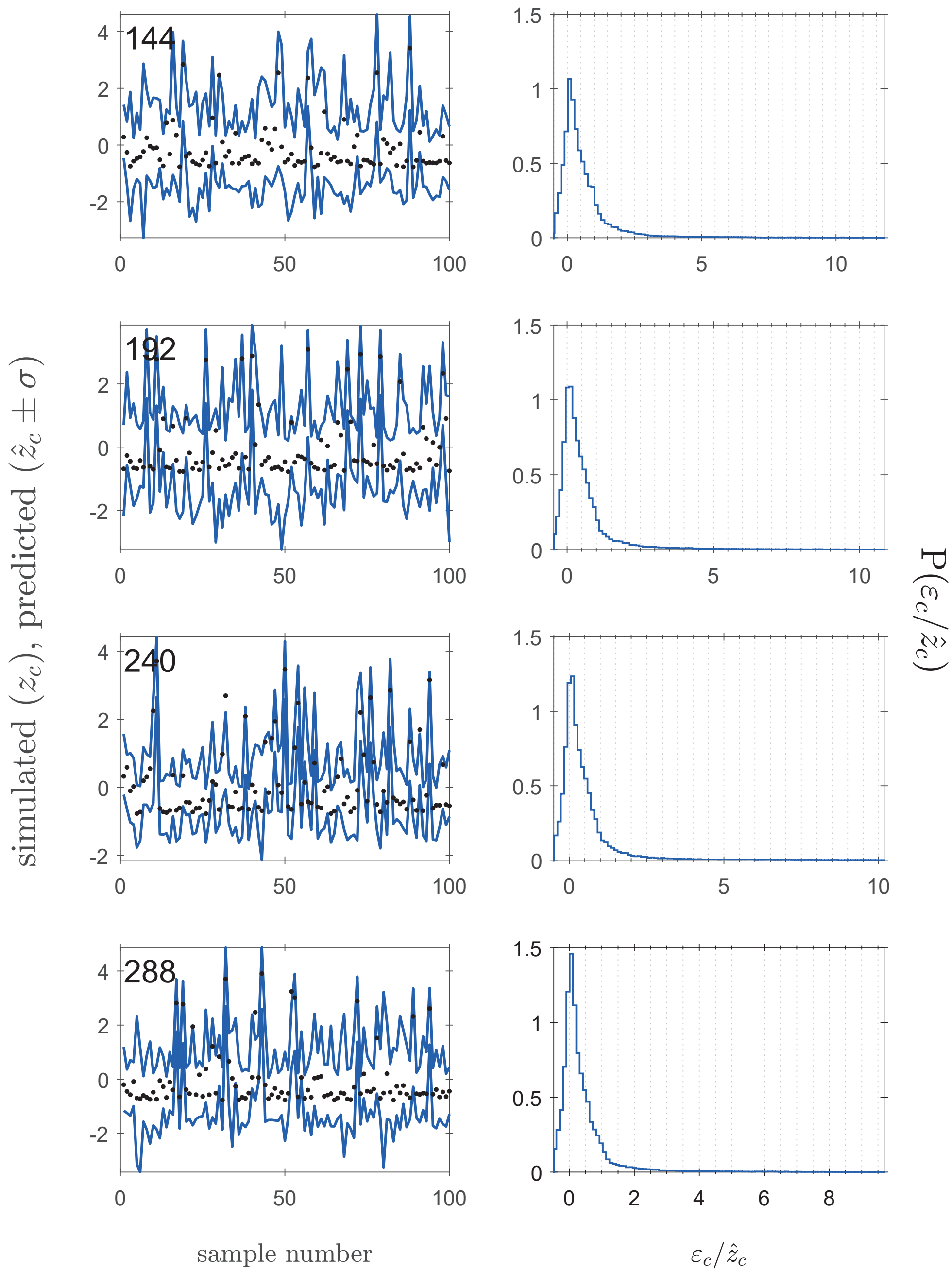


Figure 4.20 – *Case 2, Best Fit – cooling. Prediction interval (68%) of predicted values enclosing simulated values (black dots) [left]. Ratio of errors to predictions [right]. These are z-scores, not the original data.*

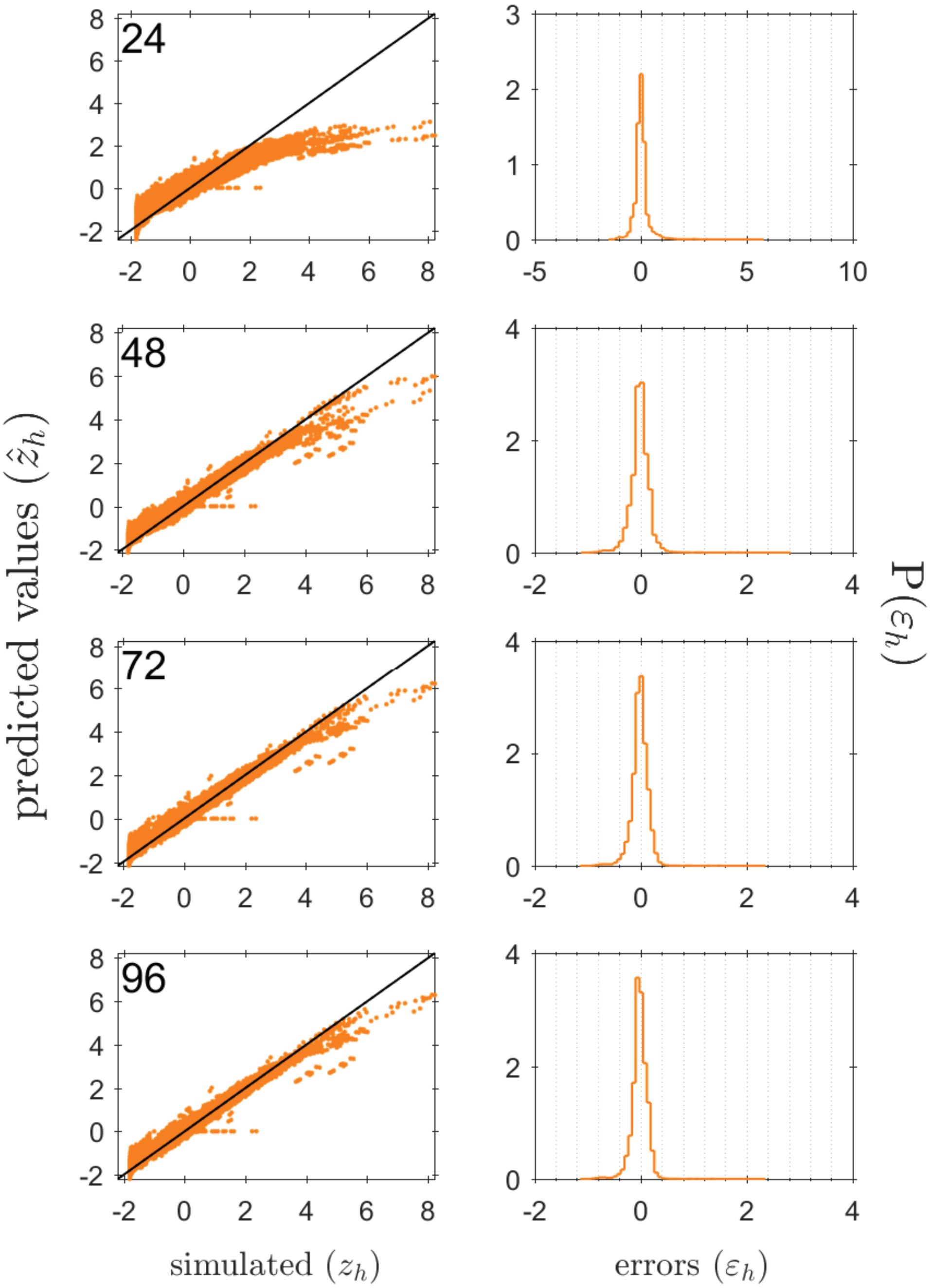


Figure 4.21 – *Case 1 (all climates), Best Fit – heating. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. Compare this plot to Figure 4.11b. These are z-scores, not the original data.*

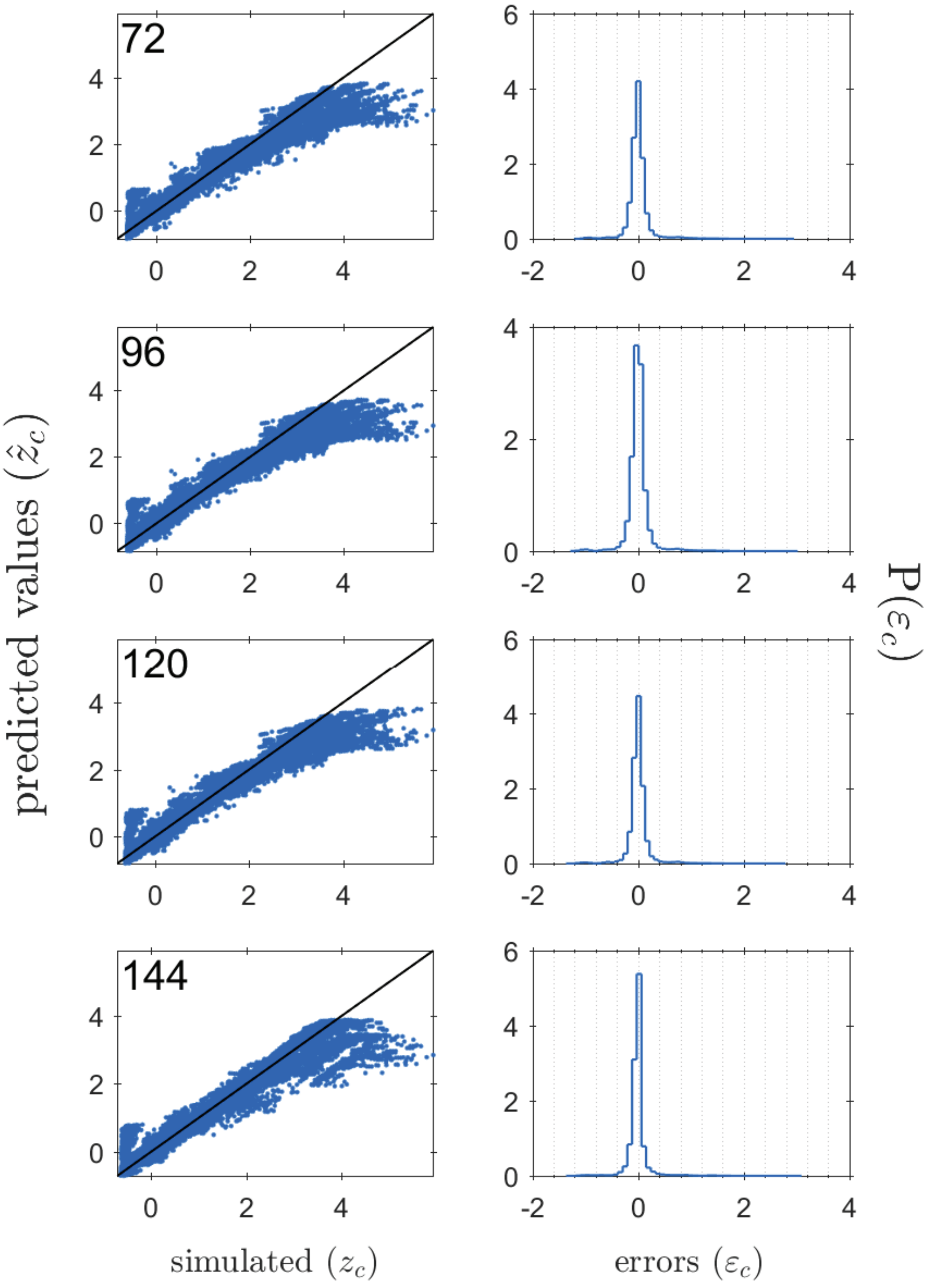


Figure 4.22 – *Case 1 (all climates), Best Fit – cooling. Predictions [left] and histograms of errors [right]. Compare this plot to Figure 4.11b, especially the number of training points used (<20). These are z-scores, not the original data.*

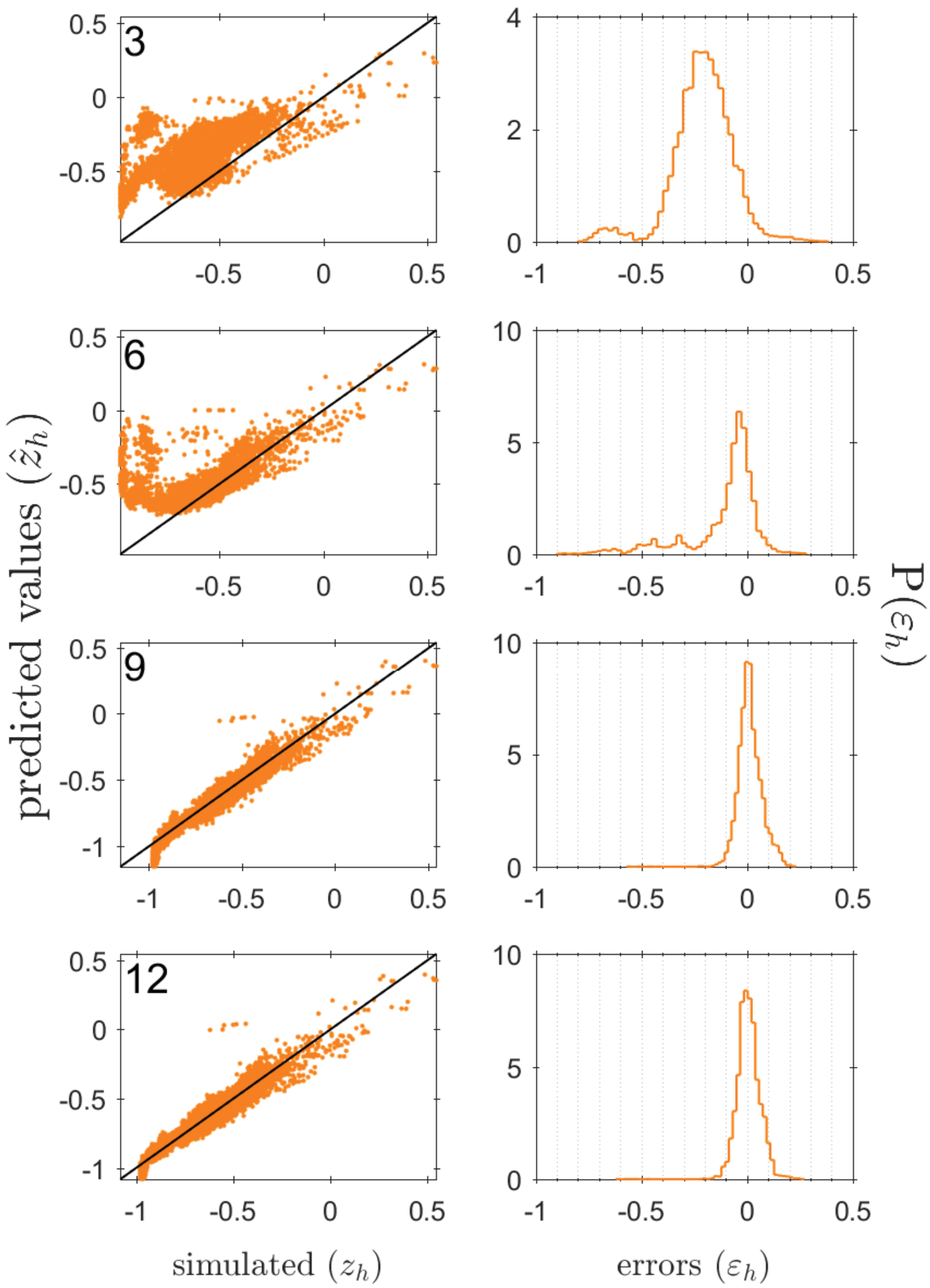


Figure 4.23 – *Case 2 (office only), Best Fit – heating. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. Compare this plot to Figure 4.11a. These are z-scores, not the original data.*

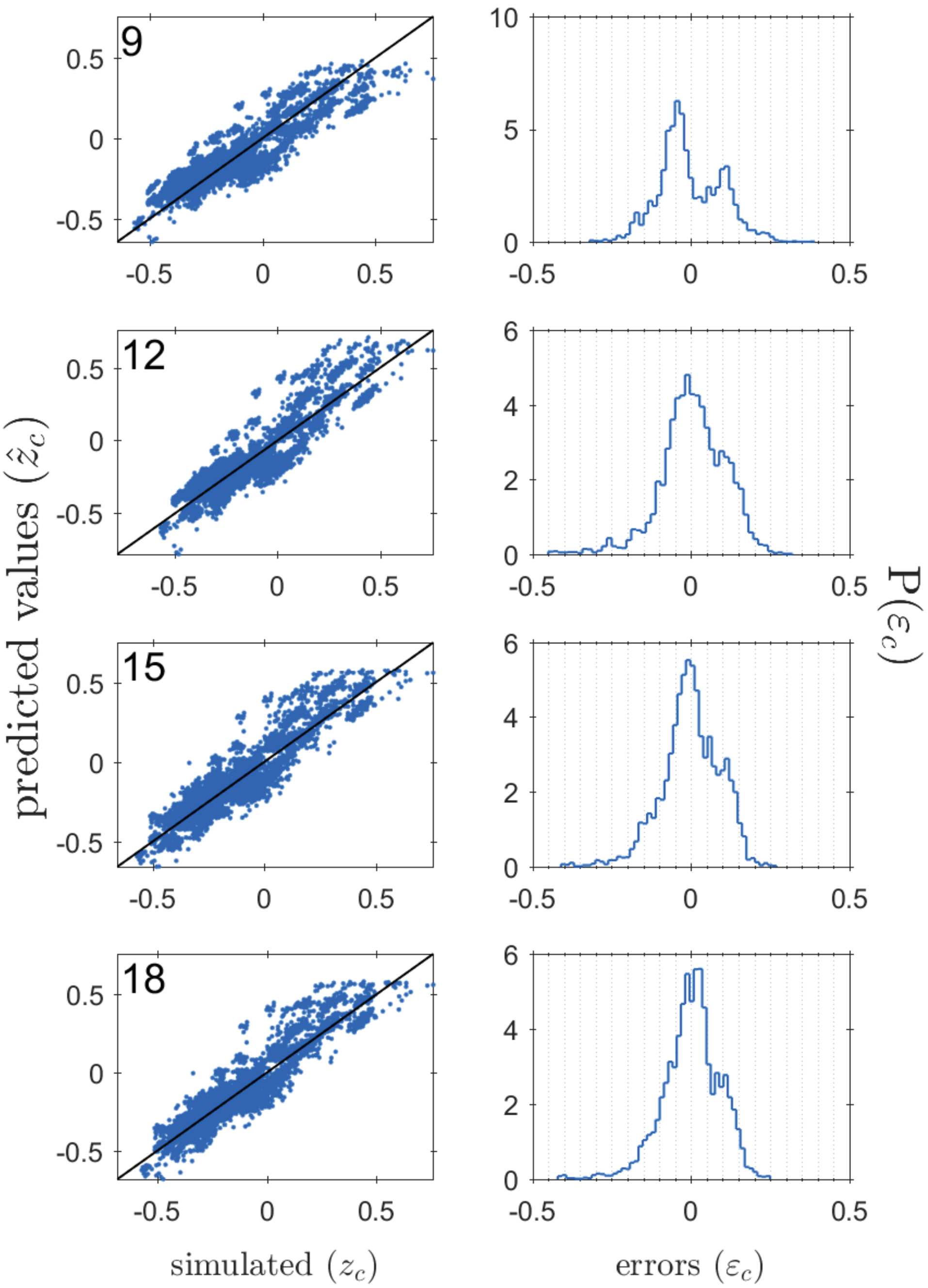


Figure 4.24 – *Case 2 (office only), Best Fit – cooling. Predictions [left] and histograms of errors [right]. See plot descriptions in Section 4.4.2.1. Compare this plot to Figure 4.11a. These are z-scores, not the original data.*

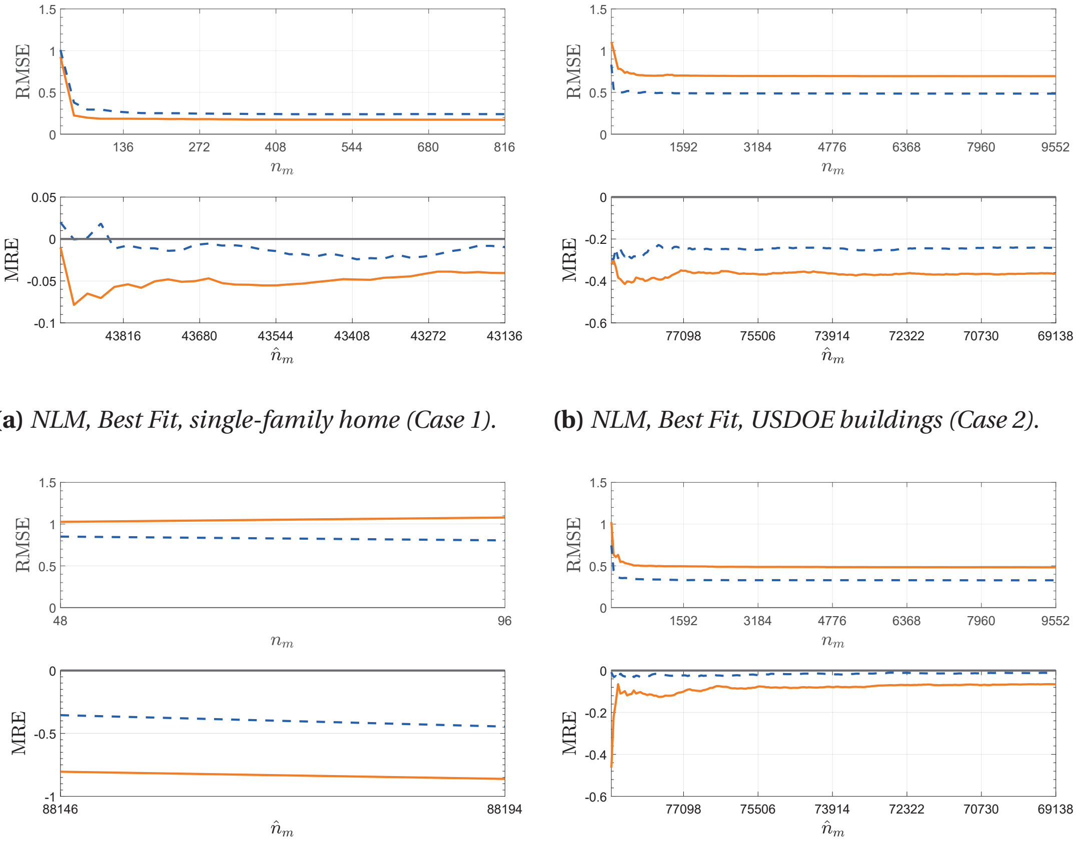
4.5 Evolution of Error: Transversal Plots

Figure 4.26 shows the error trends for the three classical model types (NLM and LMM), while Figure 4.27 shows the same for Gaussian Process regression. The RMSE is calculated using the predictions on the testing data set. As such, this is only possible in a theoretical exercise, since access to a simulated data set of the size that we have (more than forty thousand for the home, more than 78,000 for the USDOE buildings) is impractical. Presumably, in a situation where so much data is available, the simulation itself is cheap enough to not need an emulator. In this case the user is better advised to sample the original simulation wherever they please, instead of going through the trouble of fitting a meta-model. Plotted along with the RMSE is another error metric, the Median Relative Error (MRE),

$$\mathrm{MRE} = \mathrm{median}\left(\frac{\varepsilon_j}{z_j}\right), \tag{4.9}$$

where $j = 1,\ldots,n$ is an index to indicate the training data set, ϵ_j is a error, and z_j is an estimate. This is an estimate of the relative sizes of the errors compared to the predictions, analogous to a percentage error. The error metrics are plotted against the sizes of training data set n_m (top, RMSE plot), and testing data set $\hat{n}_m$ (bottom, MRE plot). The dotted line represents cooling, while the solid line is for heating. Subscript $(\cdot)_m$ is the index of models and $(\cdot)_j$ is for observations. The errors from the training sets may also interpreted as a "goodness-of-fit" measure. In each iteration (number of testing/training data points shown on the x-axis) the model is retrained with more data, with the same sampling scheme in both classical and Gaussian Process regression models (Section 4.2.5).

The general tendency of the error is a sharp drop in the first few iterations, followed by a gentle or no decline in error with more training data. Comparing Figures 4.26 and 4.27, the linear models achieve a best RMSE of about 0.2 for Case 1 and 0.5 for Case 2 (the stable RMSE value is the same for all three classical model types). In contrast, the Gaussian Process regression achieves 0.05-0.1 and 0.5 for Cases 1 and 2 respectively. The Gaussian Process regression models achieve the same errors with far fewer training data points (x-axis). The MRE sometimes starts at a very high value (close to -0.8, or -80%), but quickly reduces (as in, gets close to zero). This seems to be general tendency of the relative error – high values at the start, followed by a fast decrease. Once again, the error for Gaussian Process regression reduces much faster.



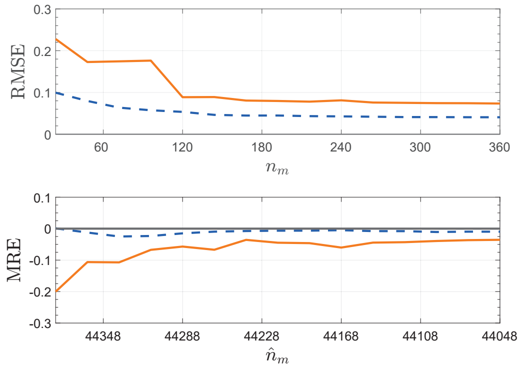
(a) *NLM, Best Fit, single-family home (Case 1).*

(b) *NLM, Best Fit, USDOE buildings (Case 2).*

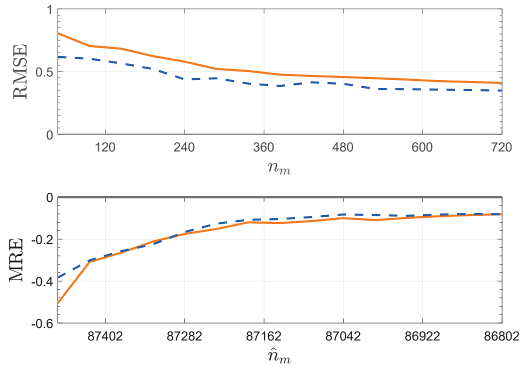
(c) *NLM, Second Fit, single-family home (Case 1).*

(d) *LMM, Best Fit, USDOE buildings (Case 2).*

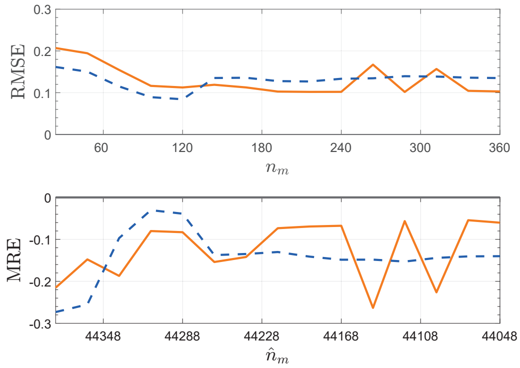
Figure 4.26 – *RMSE and MRE plots from linear models. Solid lines for heating data, dashed lines for cooling data. $\hat{n}_m$ is the size of the testing data set at each iteration, while n_m is the size of the training set. No plots are presented for LMM fit to Case 1, because the random factors (building type and climate) do not apply to that case.*



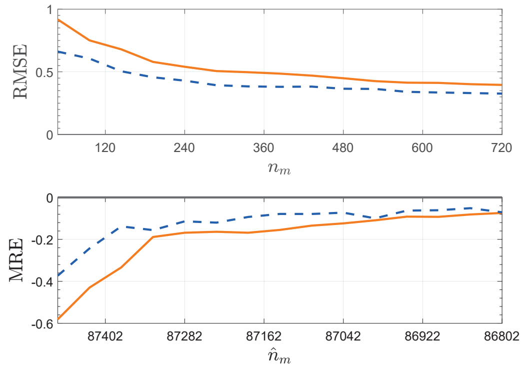
(a) *Best Fit, single-family home (Case 1).*



(b) *Best Fit, USDOE buildings (Case 2).*



(c) *Second Fit, single-family home (Case 1).*



(d) *Second Fit, USDOE buildings (Case 2).*

Figure 4.27 – *RMSE and MRE plots from Gaussian Process regression models. Solid line is for heating models, dashed line is for cooling models.*

4.6 Emulators: Summary

Faced with a choice between a theory which predicts well, but
gives us little insight into how the system works, and
one which gives us this insight, but
predicts badly,
I would choose the latter ...

R. H. Coase,
How should economists choose?
(Warren Nutter Lecture, 1981)

In this chapter, we have demonstrated the advantages of Gaussian Process regression over classical linear models in building an emulator for sensitivity and uncertainty analyses of buildings. The prediction plots of Section 4.4.2.2 are better than those of Section 4.3.3.2, with fewer training data. This means that the Gaussian Process regression models can be trained faster, i.e., with fewer simulations. In addition to better prediction on training data, Gaussian Process regression allows the performance output (energy, in this case) to be interpreted as an uncertain quantity. The downside of using this relatively novel approach, however, is that the relations between inputs and outputs are less clear than in the straightforward framework of linear models.

We are not saying, through the examples shown in this thesis, that the non-linear Gaussian Process regression approach is *always* a better choice than simpler linear approaches. In fact, a priori, we would always recommend trying linear models first. However, the expected non-linear nature of the response (comfort, energy, etc.) to most input variables (building and climate properties), makes non-linear approaches more useful. Upon comparing the results of fitting models to two different problems (Case 1 and Case 2), and modifications of the same (Figures 4.23 and 4.24 vs Figure 4.11a, Figures 4.21 and 4.22 vs Figure 4.11b), we find that the relatively more complex approach proposed in this chapter outperforms linear approaches. Gaussian Process regression trains models with less simulation points, which means that it will be easier to use for a design problem, without resorting to large scale simulation or databases. Predictions from Gaussian Process regression are also more reliable, which means that this type of regression may be used with more confidence for decision-making with uncertainty.

Textbooks on modelling and regression, e.g., Christensen (1991), Davison (2003) and McCullagh and Nelder (1983), argue that over-fitting and over-specifying a model by including a large number of predictor variables is both inefficient and does not generalise well. A model that specifies the relationship between input and output data too closely may be capturing features of the training sample that are not representative of the population. These features may be measurement errors or natural variation that should not change the mean response. The best option, and one we have repeatedly emphasised in this thesis, is to be *parsimonious*: making do with the smallest number of predictors, while explaining the training data as best as possible. The idea of parsimony is not specific to Gaussian Process regression, and is also used for the classical linear models as well as the stationary time series models in Chapter 3[28]. Parsimony does entail a loss of information, but there are neither perfect emulators nor emulators with perfect fidelity. The choice of emulators and their inputs is made using a set of principles and guidelines, many of whom are qualitative (like in Chapter 3), and have been discussed in this chapter (Section 4.2.2).

In any data-driven approach, it should generally be assumed, as is the case in this thesis, that the regression model is *incomplete*. That is, the set of inputs (covariates, independent variables) chosen are a subset of the theoretical set of all possible input variables (say $\boldsymbol{x}_{d\times 1} \in \boldsymbol{x}_{D\times 1} \mid d \leq D$). This choice is based on expert knowledge and the (empirical) results of significance tests applied to the data at hand (e.g., Section 4.2.2). Choosing a small number of covariates is better than over-complicating the model by using a large number of inputs which collectively explain the data only as well as a simpler model. It is quite possible that different data sets, i.e., those generated using different buildings or groups of buildings, may indicate that different subsets of variables are important.

The models proposed in this thesis should, thus, be used flexibly. They are designed for a specific application in mind: sensitivity and uncertainty analyses of buildings to climate, and the role of envelope properties in determining the same. By definition, the regression model can only yield information about the sensitivity of an output to the inputs that are included, and varied sufficiently, in the training data. Extending the proposed model to include other inputs of interest to a user is very feasible. It will involve training data acquired using a sampling scheme that prioritises the inputs of interest, as opposed to the ones we prioritise here (envelope-related properties). Future work, discussed in Section 5.3, mentions possible immediate extensions to include occupancy profiles, local shading, etc.

[28]Those models are also regression models, of a time series on itself, an Auto-Regressive (AR) model, or on noise, a Moving Average (MA) model.

5 Conclusion

When You and I behind the Veil are past,
Oh, but the long, long while the World shall last,
Which of our Coming and Departure heeds
As the Sea's self should heed a pebble-cast.

Omar Khayyam (ca. 1048-1122),
The Rubaiyat of Omar Khayyam, ruba'i 47.
[translator F. Scott Fitzgerald, Fifth Ed. (1889)]

5.1 Contribution

In this thesis, we have argued for the usefulness of quantifying uncertainty and sensitivity in simulation. Simulating a building with explicitly uncertain inputs specified, for example, by their probability distributions, is a practical strategy for risk-conscious design. Interpreted differently, it also implies that the design could be examined on the basis of its robustness to variation or ambiguity in an input. This work promotes the view that properly accounting for the uncertainties inherent in the inputs, only some of which are reducible with better data, is more informative than the purely deterministic procedures that are the state of the art today. At the same time, knowing the sensitivity of a design to some of its parameters allows a user to devote their resources to addressing those inputs. Thus, the most important argument of this thesis is that the *paradigm* within which building simulation is interpreted progress from a *deterministic* one that ignores uncertainty to one that *explicitly quantifies doubt*, and so allows the user to take decisions in the context of this doubt.

We have proposed procedures and algorithms for the generation of synthetic weather time series of Dry Bulb Temperature (TDB), Relative Humidity (RH), and Global Horizontal Irradiation (GHI); and, emulators based on Gaussian Process regression to make computationally-intensive *uncertainty* and *sensitivity* analyses tractable. The use of the synthetic weather time series and emulators may be interpreted as an assessment of the sensitivity a building design to changing outdoor conditions, or it may be interpreted as a quantification of the lack of knowledge about a boundary condition. While this thesis focussed on the weather input to building simulation and design, the approach of using simulation-trained emulators is easily extensible to any inputs of interest. We included, for example, envelope properties, building morphology, and internal heat gains in the emulator demonstrated in this thesis.

We interpret simulation with random inputs, or stochastic simulation, as a way to estimate confidence intervals on outputs, based either on deliberate variation of inputs (sensitivity analysis), or an estimate of the range within which an input is known with some certainty (uncertainty analysis). It is not necessary that simulation with random inputs will give a *better prediction*. In fact, strictly speaking, building simulation is not meant to be a prediction of future performance at all. If, for example, the prior distribution of an input is wrong, then the confidence intervals or variability intervals of the outputs will also be wrong. Random simulation is best interpreted, in our opinion, as an explicit quantification of what *could* be (i.e., how a building will react to a certain boundary condition). For example, it is possible that the next year has a very warm and long summer, so cooling loads (or overheating) could be several times what was seen in a typical year. Alternatively, a mild winter could be followed by a cold spring, stretching and damping fuel usage in the northern hemisphere. As we have argued before, both of these are simultaneously possible in the present, but a user has no way of knowing which scenario will eventually happen.

The ideas on which this work is based – synthetic weather and emulators for quantifying uncertainty and sensitivity – have certainly not been discussed here for the first time. We show examples of uncertainty and sensitivity analyses using external and internal methods (section 2.4), and the use of emulators for computationally-intractable simulation (section 2.5). We also discussed the state of the art and history of the generation of synthetic weather series (section 2.6.3), and climate classification (section 2.1). The novelty of this thesis, thus, is in the two separate but related ensembles of methods that enable rapid, practical, and reliable sensitivity and uncertainty analyses of building simulation outputs. We address the challenges of computational load, complexity, feasibility, and generalisability identified in chapter 2, which shave been significant barriers to the widespread adoption of uncertainty and sensitivity analysis, especially vis-à-vis weather. The use of time series models to create variations on such small source data sets (only one 'year') has been demonstrated for the first time in this thesis. The use of Gaussian Process regression to create emulators with such a small number of initial 'training' simulations, such that a custom emulator can be created for every design problem, has been demonstrated here for the first time.

The procedures proposed in this thesis are meant to be used in tandem with building simulation. In the following section (Section 5.2), we lay out a schematic of how the procedures may be implemented (following Figure 5.1). As mentioned in Chapters 3 and 4, the computer code/scripts implementing our work may be found with the archive copy of this thesis on infoscience.epfl.ch. Section 5.3 discuss the applications and limits of the methods proposed in this thesis, the work necessary to realise these methods as tools for widespread use, and applications in contexts not discussed in this thesis. Limitations and improvements for the modelling procedures described in this thesis have been discussed throughout the main text chapters 3 and 4. In this section, we focus on the uptake of the methods and their envisioned impact on practitioners and clients. These are given special mention here because they affect the relationships this work will have with the larger field of Building Performance Simulation (BPS).

PRÉCIS

- All inputs to building simulation are uncertain, and different inputs may be known with differing levels of confidence, for a variety of reasons.
- Not all of this uncertainty can be *eliminated*, though some of it can be reduced.
- A solution/design focussed on just the mean or deterministic inputs only answers the requirements under these mean conditions.
- The actual conditions experienced by the building, and its as-built properties, may vary substantially from the mean.

■ Explicitly including the uncertainty of inputs through, for example, approximate confidence intervals constructed through random simulation, may improve the robustness of design by calculating, and designing for, the variations in boundary conditions. However, this *does not* imply an improved prediction, since even a stochastic input (to create a variability interval through Monte Carlo Analysis (MCA)) could be wrong.

5.2 Workflow Summary

Referring to Figure 5.1, the workflow is described below. This is a summary of the mechanics of using this work, while Section 5.3 describes where and when it may be used.

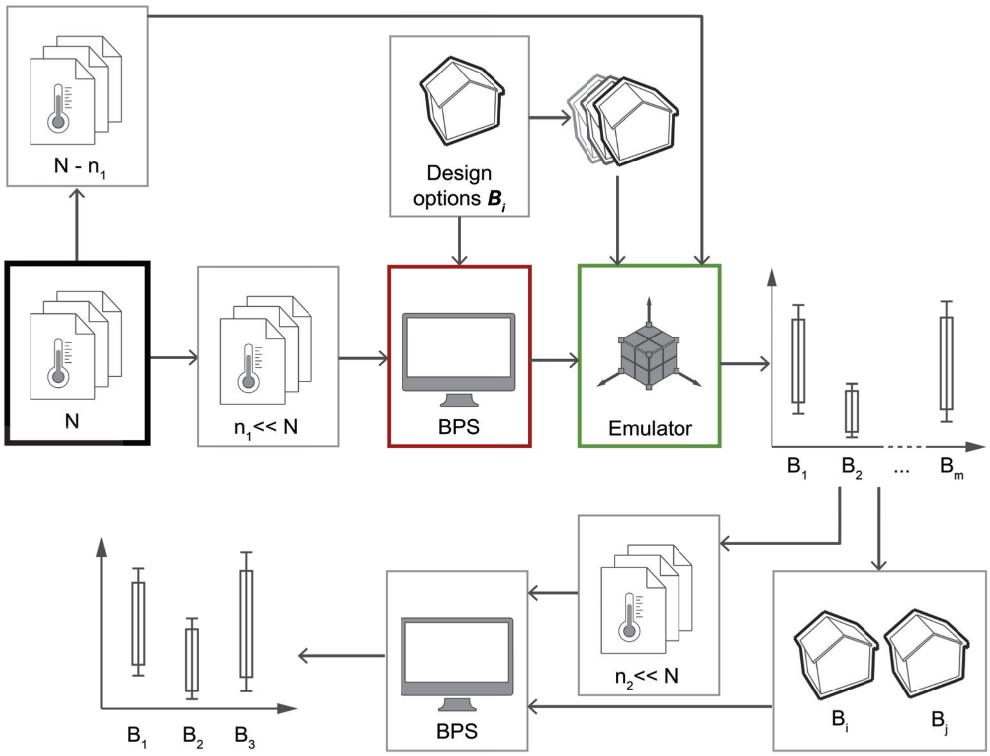


Figure 5.1 – *A schematic of the envisioned workflow.*

1. Use the procedure described in Chapter 3 to generate synthetic files, with or without climate change forecasts included, from a typical weather file (N weather files, usually 100-200). The output of this step is a number of synthetic weather time series (e.g., Figures 3.16 and 3.25).
2. Model the building being studied in any energy modelling software. Simulate the building model with a few weather files (10-20, $n_1 << N$ in Figure 5.1). If some design options or renovation strategies are known (B_i), model those and include them in the simulation. The outputs of these simulations may be plotted as in Figure 5.2 or Figure 4.4.
3. Extract the building properties mentioned in Table 4.1 from the simulation res-

ults (e.g., U-value), or any other properties of interest. The building properties used in this thesis mostly relate to the building envelope, but the user can work with other design parameters according to their interest (e.g., building compactness, ventilation levels). These properties serve as independent variables for the regression model (emulator). The properties are either calculated by the user (e.g., Window-to-Wall Ratio) or may be obtained from simulation (e.g., annual average self-shaded fraction of the facade).

4. Calculate the climate-related properties for each weather file input (e.g., median TDB), as described in Chapter 4 and appendix B. These are also independent variables for the regression model and involve only simple arithmetic (e.g., median, sum).

5. Train a regression model using the procedures described in Chapter 4, where the building- and climate-related properties extracted from the initial simulations are the independent variables. The dependent variable of this regression model is either heating or cooling energy need/usage (or other quantities of interest like peak demand). This regression model is now an 'emulation' of building simulation, and may be used in its stead.

6. If the objective is to explore various refurbishment/design options, they should be expressed as changes in the building properties which have been used as inputs to the regression model (e.g., change in U-value to test a high insulation scenario). The climate-based inputs (independent variables) should not be changed in this use case, since the climate is a 'boundary conditions'. Rather, at each combination of building-based independent variables, i.e., one design/refurbishment option, the emulator should be evaluated with a wide range of climate-based inputs. This gives a range of possible future energy outcomes for a given building configuration (or design, or combination of properties). This would enable to user to make a comparison such as that given in Figure 5.2. If Gaussian Process regression is used, then the output (e.g., heating energy usage) at each query point (combination of building- and climate-based inputs) is a random variable with some mean value and variability.

7. If the objective is to explore the impact of the uncertainty due to climate, or the sensitivity of a building design to climate, or the expected interior conditions under some specific external conditions (e.g., heat wave), use only specific levels of the climate-based inputs (i.e., climate-based properties calculated from specific files) for any design/refurbishment option.

5.3 Scope and Applications

… this too is my instruction:
proceed along paths which wagons do not traverse,
and do not drive along the same tracks as others nor on the broad highway,
but along [fresh] ways, even if your course will be narrower …

Callimachus, fr. I, vv. 1— 7, 17— 28[1]

5.3.1 Random Simulation as Design-Assistance

The point of building simulation, ultimately, is to assist planners and designers to improve designs (from the perspective of comfort and energy), allocate resources, and anticipate demands. Therefore, the extra effort of uncertainty or sensitivity quantification is only justified if it enables planners and designers to make better decisions. We envision that with improving computational access and familiarity of designers and clients with quantitative decision-making, the use of confidence intervals, variability intervals, and other measures of variability and uncertainty will become ubiquitous.

In the scope of this thesis, we did not explicitly test the proposed methods in a design setting. That is, we did not test whether designers make a 'better' decision if they are aware of the uncertainty of their calculations. This is not exclusively a question about this thesis, but rather about whether professionals make better decisions if estimates of uncertainty or dispersion are available. It is not clear that providing designers with these estimates of variability will necessarily change the decision they would make based on mean values. It is often difficult, in any case, to arrive at a consensus on what a better decision is. In some cases forensic analysis (e.g., through post-occupancy evaluation) may help. In this thesis, we define 'better' as being more robust to climate volatility. In other contexts, it could also be defined as, for example, a decision that minimises the risk of failure, at the cost of higher initial capital expenditure for more larger mechanical systems; or a decision that aims for the lowest operating energy in the majority of anticipated climate conditions, but allows for failure; etc.

[1] Cited in the Introduction to *Jason and the Golden Fleece (The Argonautica)*, from Apollonius of Rhodes, translated by Richard Hunter (Oxford World's Classics, pg. xvii).

Table 5.1 – *The renovation cases of a single-family home. This table is a reduced version of Table B.4.*

Old Code	New Code	U-value	Description
BC00	G000C00	1.70	Base Case
–	G000C01[a]	2.65	
–	G000C02[b]	2.85	
RC01	G000C03	1.09	External wall insulation
RC02	G000C04	1.04	
RC03	G000C05	0.98	
RC22	G000C24	1.70	Envelope airtightness

[a] Masonry-only wall
[b] Brick-only wall

5.3.2 Renovation Strategies and the Performance Gap

In addition to the potential of random simulation to impact design decision-making, we expect that sensitivity and uncertainty analysis may help to address the so-called 'performance gap' in energy-conscious design and refurbishment[2]. This is not to say that the predictions will be more 'correct', because random simulation cannot by itself *reveal* future conditions. Rather, it will allow the designer and client to discuss performance in terms of confidence and risk, bringing expectations closer to delivery. This thesis proposes, for the first time, the ability to quantify this risk due to climate, and lays the groundwork for similar studies for other uncertain inputs (such as occupancy, or construction variation), through the use of probabilistic emulators.

As discussed before, we propose that building simulation always be carried out in a stochastic paradigm, where design options or interventions are always assessed keeping in mind their robustness (expressed through a spread of energy use) to a volatile climate. The argument for why this kind of analysis is important is included in Chapters 1 and 2. An example of how different design options may be evaluated through a comparison of their energy use spread is in our previous work (Agarwal, Rastogi et al. 2016; Chinazzo 2014; Chinazzo, Rastogi et al. 2015a,b).

[2]Energy-conscious buildings often consume far more, and occasionally less, energy than calculated from simulation. This gap is understandable from the point of view of the expert simulator, and is generally attributable to usage, construction errors, and unplanned changes (Gupta and Dantsiou 2013). It may, however, cause 'sticker-shock' for the client if the building uses, say, *twice* as much energy as planned, and undermine the credibility of the energy consultants and designers.

Examine the comparison of refurbishment options shown in Figure 5.2 and table 5.1. In this case, the decision to install wall insulation (options 03-05), for example, is significantly better than doing nothing (option 00). The improvement of infiltration (option 24) is by far the best option, both in terms of robustness (spread of box plot) and median energy use. However, if the user is deciding between the three levels of insulation, the choice is less clear. While it is true that more insulation results in lower heating energy usage for each weather file, all the files are equally probable, so the future predicted savings will be very different from what would be predicted by using just a single typical file. The point is not that more insulation will ever result in more energy use, rather it is to say that the payback from adding more insulation is not the same as that would be predicted by using a single simulation (typical file), and then assuming that number to be representative of all future years.

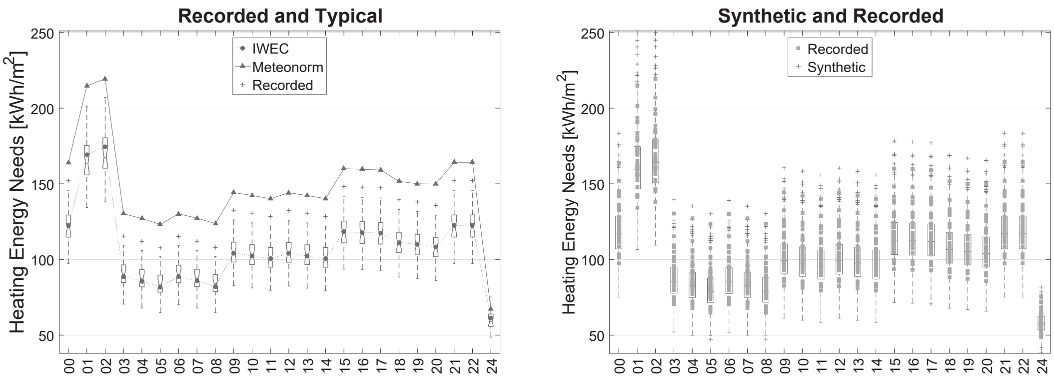


Figure 5.2 – *Heating energy use by twenty four design options/renovations for a single-family home, including the current construction (option 00). On the left are the same simulations presented in Figure 1.1, the plot on the right includes synthetic weather data as well. While these boxplots have been produced using EnergyPlus simulations, they may also be produced with an emulator.*

5.3.3 Comfort Assessment

An important application of this work is in the assessment of future indoor conditions without mechanical systems. We envision the calculation of both risk, e.g., summer overheating, prolonged high temperatures, etc., and the change of 'normal' indoor conditions, e.g., due to a change in summer median temperatures over the next century. This is most often examined in the context of summer overheating in buildings without cooling systems (e.g., Agarwal, Rastogi et al. (2016)). Testing buildings' response to extreme events like heat waves and/or a slowly shifting climate has hitherto been limited to a few specialised case studies (see Chapter 2 for a discussion of the state of the art). In addition, the literature offers only single files for future studies, the potential pitfalls of which we have discussed in Chapters 1 and 2. The methods developed in this thesis offer multiple pathways to assessing risk – whether through the simulation of a large number of synthetic weather years to obtain distributions, or the use of extreme files to obtain limits.

A limitation of the work presented in this thesis is that comfort is treated as a rigid concept: we use fixed set-points for heating and cooling. Since comfort was not a focus of this thesis, we focussed on ideal loads derived from fixed set-points to demonstrate the emulators. An ideal load calculation does not consider a system, so the indoor conditions are examined in view of their (energy) 'distance' from some arbitrary comfort threshold for indoor temperature and humidity[3]. This is a choice that we leave open to the user, since it depends on choices such as whether the building is conditioned or free-running. Future work to implement the proposals of this thesis should include the testing of the emulator's utility with adaptive models (usually applied to free-running buildings). Given the data-driven nature of regression-based emulators, this change is unlikely to affect their ability to explain the data. However, adaptive comfort does introduce an additional, potentially non-linear, factor which demands detailed examination. That is, it would be yet another reason why the uncertainty of inputs will not propagate linearly/additively. Introducing adaptive comfort would entail a significant reduction in loads for moderate climates, i.e., there will be far more hours in a year without a need for heating or cooling. In the simplest scenario, this would merely shift the distributions presented in figs. 4.4a and 4.4b, reducing their mean but not changing their shape. In a more complicated scenario, the shape of the distributions would change entirely. Combining the concept of comfort with hourly emulators of building response, the loads could be modelled as a combination of a Markov chain (load vs no load) and a magnitude function (how much load), akin to the weather generators described in section 2.6.3. This would involve the crossing of time series models with emulators, similar to the work described in section 5.3.7.

5.3.4 Model Predictive Controls

Predictive building automation systems are based on modelling the state of the building a few time steps in the future. This is particularly important for buildings with high thermal mass, and therefore a slow thermal reaction, since actions taken at a particular

[3]We discussed the idea of an enthalpic 'distance' on a psychrometric chart that must be 'traversed' for a certain indoor condition, created passively by the building, to reach some comfort limits in Rastogi, Horn et al. (2013).

time will have consequences well into subsequent time steps. The synthetic weather data proposed in this thesis could be used to train these models before they are deployed, and for testing the suitability of different control algorithms. The advantage of using synthetic weather files is the same as that of using them for uncertainty and sensitivity analyses: much larger quantities of plausible data for a particular location than would be available from records alone. Several control algorithms available today ‘learn’ user behaviour or preference, e.g., Lindelöf (2007), and the learning algorithm could be trained on the large quantities of synthetic data that is readily available from the use of our generator. Another perspective on this would be to design model predictive controls to explicitly account for the uncertainty in weather/occupancy predictions, and plan to maintain indoor conditions within some boundaries, even when the weather/usage comes close to some arbitrary value (e.g., twice as much occupancy as planned).

5.3.5 Modelling Stocks, Grids, and Renewables

We have alluded to the idea of extending the use of emulators for rapid sensitivity analyses, what-if studies, and climate risk assessment of groups of buildings, or building stock (section 2.1). The problem of simulation time is more acute when modelling large groups of buildings, or neighbourhoods. We have included some factors that may be useful when modelling groups, like form factor and roof ratio, building on the work of Nault (2016), Nault, Peronato et al. (2015) and Nault, Rastogi et al. (2015). We are certain that more factors will have to incorporated, perhaps including site-specific factors such as shading among buildings[4], or access (to the sun). A straightforward application of the emulator and synthetic weather would be the simulation of random, weather-driven, electricity or gas demand from groups of buildings.

This thesis addresses the ‘demand’ side of the energy balance, i.e., buildings, and may be easily extended into the ‘supply’ side, electrical grids. This could be done for two applications: simulating operations based on random weather, and assessing risk in future production values. The sensitivity of renewable energy production, especially solar power, to climate could easily be carried out with the stochastic weather generator proposed in this thesis because weather is the primary driver of (potential) renewable production. In ongoing work, we are collaborating with economists to assess the risks in solar investments based on climatic uncertainties. Similar to the training of model predictive controls for buildings (section 5.3.4), the control or operation of electric grids could also be simulated with a combination of

[4] In addition to the self-shading parameter, average sunlit percentage, already used in this thesis

random weather and random occupancy feeding a rapid-response building emulator.

5.3.6 Stochastic or Robust Optimisation with Uncertain Inputs

Stochastic or robust optimisation would, in this context, imply optimisation that explicitly includes the uncertainty of inputs (in the form of a distribution) *during* the calculation of a cost function. One way of doing this, which we provisionally label 'static optimisation', would be to run an optimisation routine using several different weather files one after the other. Then, one would have a distribution of the optimal values of, say, insulation level. This approach has been used by, for example, Ramallo-González, Blight et al. (2015).

This could be carried out using an emulator based on Gaussian Process regression by fitting the regression function to uncertain inputs. Recall that, unlike kriging, Gaussian Process regression does not need to assume that the training data points are exact (see section 4.4.1). The advantage of this approach would be a robust design solution, that is 'optimal' for a range of operating conditions (weather, occupancy, construction errors, etc.)

5.3.7 Regression of Internal Temperatures

For a brief introduction to regression-based time series models, see Cryer and Chan (2008). Regression is a technique that pervades most of this thesis, since the time series models are also *regressions of a time series on itself*. In this future application, however, we propose the use of a different flavour of regression with time series.

The results of our simulations for this thesis point to the influence of external weather parameters on the internal temperature. So far we have been using an ideal loads system with infinite capacity. A more realistic case for a large proportion of dwellings in Geneva, for example, is a heating-only system. This means that in the summer, these buildings would have a 'free-floating' internal temperature, creating significant overheating risk. Oraiopoulos, Kane et al. (2015a,b) model internal temperature as a time series to predict overheating risk. In their case, the application is to short term forecasts (2-4 days) to create an early-warning system for at-risk dwellings in the UK. Their approach is to use Auto-Regressive Moving Average (ARMA) models as well, but it is based entirely on external weather conditions. We propose a future study that is very similar, but with additional parameters to account for the building construction, particularly its envelope. We feel that this is a crucial component for the prediction of internal temperature since the response of a building is a function of its construction, an argument we have also made in chapter 4. The exploration of time series models that account for building construction in their lags and coefficients could be a useful tool for modelling overheating risk in buildings, both for annual figures and short-term forecasts.

5.3.8 Random Occupancy Modelling

Occupancy modelling has been previously examined in terms of both uncertainty and sensitivity analyses. See, for example, Ramallo-González, Blight et al. (2015) for robust optimisation considering building behaviour; Haldi (2010, 2013) for a Markov chain model of adaptive actions for thermal comfort; Mavrogianni, Davies et al. (2014) for the interaction of occupancy patterns with the risk of overheating in London dwellings; among others. In the emulator demonstrated in this thesis, occupancy profiles have not been examined in detail, having been represented merely as fixed Internal Heat Gain values. A study of the interaction of random or varying occupancy and random weather is planned. This study is envisioned at two levels: one where the interaction of the two kinds of random profiles is purely coincidental, and one where occupancy responds to weather. In the first type of study, the occupants' schedule and actions may be thought of as independent of the weather. In the second type, we would use previous research, e.g., Haldi (2010), to model the adaptive actions of occupants based on the weather and indoor conditions. The interaction of weather and occupancy is, we feel, crucial to future predictions of overheating risk and the robust design of buildings in a changing climate.

5.4 Outlook

... the future's uncertain and the end is always near.

Jim Morrison, The Doors
Roadhouse Blues (Morrison Hotel)

Upon reading this thesis, it is not unreasonable to conclude that predicting *exact* building energy consumption at a *specific* time in the future is futile. Even if the software representation of a building is completely reconciled to the design, or as-built conditions, neither future weather conditions nor occupancy and usage are exactly predictable. These two inputs substantially affect the reliability of simulation results, even if the effect of an unsuitable value for any of a number of parameters and settings is ignored. Essentially, the calculations, inputs, and parameters in simulation are all approximations of natural and human factors about which we cannot have complete knowledge. This conclusion, though, is far from fatalistic, since we are calling for a reinterpretation and reformulation of existing practice with new tools, not the abandonment of simulation in design. Discussing building simulation in terms of uncertainty, confidence, and risk does not make it less useful; on the contrary, these considerations make the exercise *more* informative. The proposals contained in this thesis are not intended as ironclad rules, and our approach implicitly counsels against rigid interpretations in favour of flexible, risk-aware decision making.

We are not in favour of the so-called ‘look-up table’ approach, where a database of simulations is used to approximate the energy performance of any new case being studied. Databases of measured energy and other metrics do have a role in improving building energy estimation (through the use of benchmarking, for example). However, a method that deliberately blurs the uniqueness of each design problem by shoehorning projects into pre-defined types and scenarios seems unnecessarily simplistic. Regression is also a simplification, and is conceptually very close to a look-up table, but these tables are deterministic whereas the regression-based approaches we propose are not. If the boundary conditions of a problem are inherently random and uncertain, then it is more appropriate to work with methods that allow this randomness to be explicitly stated.

A persistent concern with the ideas and procedures described in this thesis is their obvious complexity when compared with the current practice of working with single weather files and fixed building inputs. As we have expressed elsewhere in this thesis, while using stochastic approaches is indeed more time-consuming, we question the appropriateness of using deterministic inputs to represent random phenomena. We anticipate that the use of uncertainty-based approaches requires a change of mindset, where consultants/designers do not deliver precise performance figures (e.g., future annual energy use), and clients do not expect it. While we have not, in this thesis, examined the understanding of uncertainty by designers and users, we expect that the concepts are not necessarily simple to integrate into the everyday workflow; particularly since simulation is taught in the “deterministic paradigm” we describe in this work. Our work, then, both relies on a change of mindset and proposes new techniques that will facilitate this change.

A common concern with incorporating any new technique in the simulation workflow is its benefits compared to the costs and complexity. We feel two particular issues should be answered: firstly, given that systems are usually over-designed, will the inclusion of random inputs necessarily change the calculations; and secondly, given the complexity of both conducting and interpreting simulation inputs and outputs, will it not be necessary to completely rethink the standards and rules that have been codified to regulate its use in design? We offer broadly optimistic responses to both concerns. System over-design is understandable from both liability and robustness perspectives, and in fact passive buildings with smaller systems will demand more adjustment from occupants. We do not take a position on how much adaptation should be taken for granted and in which situations. Rather, our work offers the tools to calculate this and other climate-based risks systematically and easily. While we have not explicitly addressed the impact of uncertainty analysis on regulations, the experience of other professions offers a guide. The financial and actuarial industries have a long record of using risk assessments. Most professional bodies anyway update regulations and codes regularly. We contend that explicitly requiring the systematic consideration of the uncertainty in inputs is both feasible and necessary. Ultimately, however, neither regulations nor theses can bring about the change of mindset that is required to adopt random simulation. It is more likely that the adoption of statistical training and an appreciation of robust simulation practices in university curricula will bring about lasting change in practice.

The arguments and proposals of this thesis are a first step in bridging the gap between the benefits of so-called "big data" and the human ability to process it or use it for decision-making. We offer some specific proposals in this chapter, and this would be one of the natural extensions of this work. The influence of large datasets on building design and operation has has been curiously muted. Not all of the blame for this lies necessarily with the building-related trades, since the industry is both conservative and fragmented, for natural reasons, and new trends and techniques may take a generation to be accepted. However, the uptake of computer-aided tools is now nearly universal, and designers and engineers are increasingly comfortable interacting with intelligent programs in their day-to-day work. Eschewing hard optimisation, we have argued for an expert-system style approach, updated and informed both by random simulation and measured data (if available). The inclusion of large-scale data processing, whether based on simulation like the proposals of this thesis or the use of measured data, have a good case for changing the way we think about designing buildings for evolving expectations in a changing climate.

We end with a prophetic observation from Pliny the Elder:

. . . this alone is certain,
that there is nothing certain . . .

Pliny the Elder, The Natural History
Book II, Section 7/Chapter 5.
Translated by John Bostock, perseus.tufts.edu

A Additional Results and Concepts: Synthetic Weather

गुलों में रंग भरे बाद-ए-नौ-बहार चले
चले भी आओ कि गुलशन का कारोबार चले ।।

फ़ैज़ अहमद फ़ैज़
गुलों में रंग भरे

Filling flowers with colour
comes the breeze of new Spring
O! Do come [dear Spring],
so the gardens may bloom again.

Faiz Ahmed Faiz,
Filling Flowers With Colour

A.1 Notes on Implementation

We tested the synthetic values against recorded data obtained from the National Climatic Data Center (NCDC)[1] and various national meteorological agencies, the full list of which is in Section A.5. The synthetic data are created entirely from the commonly-used typical year files. These files may be obtained from the web site of the Energy Plus software[2], the *Climate.OneBuilding.Org* website or from commercial suppliers such as METEONORM (MN)[3] or *White Box Technologies*[4]. The output from climate change models is introduced as a replacement to low frequency Fourier fits in this chapter (Figure 3.5). The climate change model outputs were obtained from the WCRP CORDEX website [5].

[1] http://www7.ncdc.noaa.gov/CDO/cdo
[2] https://energyplus.net/weather
[3] http://meteonorm.com/download/software/mn70/
[4] http://weather.whiteboxtechnologies.com/home
[5] http://www.cordex.org/

The procedures/data are based on several other climates besides the three presented here (full list in Tables B.5 to B.8). For additional climates and results, as well as original code, the reader is directed to the supplementary documents provided with the archival copy of this thesis on the EPFL[6] repository (infoscience.epfl.ch). The text of this part of the thesis builds upon work previously published in Rastogi and Andersen (2013, 2015, 2016). Figures 3.1 and 3.5 were created with Siobhan Rockcastle. The climate change forecasts from the CORDEX website were downloaded and processed by Georgios Mavromatidis.

A.2 Fourier Series and Fitting

A *generalised Fourier series* is a series expansion that uses a complete orthogonal system of functions to represent any arbitrary function. We will be using the discrete analogue of a generalised Fourier series using the common trigonometric functions of sine and cosine, which are orthogonal over $[-\pi, \pi]$. In this application we are not fitting the Fourier series directly to a function $f(x)$, but instead to a finite data set y_t which is generated by some unknown underlying function. A Fourier series representation with infinite sine-cosine pairs does not need an error term. If the number of trigonometric pairs m is equal to half the number of data points, $\frac{n-1}{2}$, the number of independent variables is n, the same as the number of time steps or data points. This implies a perfect fit to the data because such a model is *saturated* (Christensen 1991, sec IV.2).

Fitting a Fourier series to a sequence of observed values begins with the identification of the principal harmonics, or the frequencies of the dominant waves. A handy way of doing this is through a periodogram, shown in Figure A.2. A periodogram is a plot of the estimated Power Spectral Density (PSD), at frequencies of interest, of a "wide-sense stationary process" (The MathWorks, Inc. 2015). The quantity plotted is the amplitude of each trigonometric pair, i.e., the mean sum of squares at any frequency. It can be obtained by taking a Discrete Fourier Transform (DFT) of the time series, and it gives an indication of the 'importance' of a particular frequency. Periodograms are of limited value if the data in question is not a stationary time series, since the PSD at the *0^{th}* frequency or the mean will practically drown out any other harmonics of interest. In Figure A.2, we have deliberately not plotted the power at the 0^{th} frequency.

[6] École polytechnique fédérale de Lausanne, Switzerland.

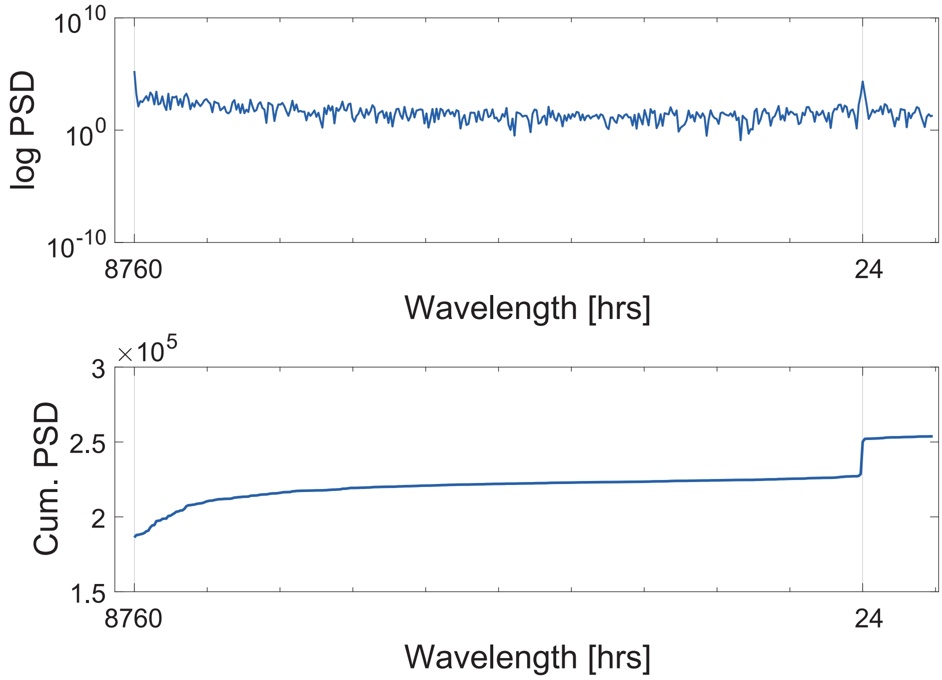


Figure A.2 – *Periodogram of the Dry Bulb Temperature (TDB) time series for Geneva. Note that the y-axes use log values. The x-axis is labelled with the wavelengths rather than the frequencies.*

A.3 Stationary Time Series Models

Christensen (1991) describes "time domain models" as "linear filters of the white noise process". The most general form of a stationary time series is a *general linear process*

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots, \tag{A.1}$$

or,

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{A.2}$$

where, ε_i are white noise samples[7], and ψ_i are weights. It is essential that $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Equation (A.1) is a "weighted linear combination of present and past white noise terms" (Cryer and Chan 2008). Modifying Equation (A.2) gives *Wold's decomposition* (Wold 1938)

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} + H_t, \tag{A.3}$$

where, ε_{t-i} is an (uncorrelated) white noise sequence that is the input innovation process to y_t via the linear filter ψ_i, and H_t is a deterministic term, which is perfectly predictable based on past values of y_t. The historical information (past values) could be on one or more of the following types: **past values** of the process itself, uncorrelated white noise or **innovations**, and past values of some exogenous variables. Exogenous variables are those that are considered to be unaffected by the process y_t. Innovations and past values of a process are used in this thesis, while the use of exogenous models for time series regression is proposed in future work. The models we use are *causal*, i.e., they predict the value at time t based on past values up to that point. One can see that Equation (A.3) contains the seeds of an Auto-Regressive Moving Average (ARMA) model (see Section 3.5), as it permits the modelling of a variable y_t based on a linear combination of its own past values and an independent noise sequence.

Changing the deterministic term to be the history of a variable up to the preceding time-step, H_{t-1}, and taking expected values, gives a *conditional mean model*

$$E(y_t | H_{t-1}) = \sum_{i=0}^{\infty} (\psi_i), \tag{A.4}$$

where $E(y_t|H_{t-1})$ is the expected value of y_t conditional on the preceding history of the process H_{t-1}. The unconditional mean of the stationary process, $E(y_t) = \mu = 0$, so it is not mentioned here. In other words, Equation (A.4) is a calculation of the conditional expected values of the series in Equation (A.2). The two models we use in this thesis, Moving Average (MA) and Auto-Regressive (AR) models, are specialised forms of conditional mean models, fitted to linear processes with Gaussian innovations. An ARMA model is simply a combination of these two.

[7] A random signal with a constant PSD.

A.3.1 Auto-Regressive (AR) Models

In models where the historical information H_{t-1} consists solely of the past p values of a variable, we get an $AR(p)$ model

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{A.5}$$

or, in lag operator polynomial notation,

$$\phi(L) y_t = c + \varepsilon_t, \tag{A.6}$$

where, $\phi(L)$ is the lag operator polynomial representing the linear combination of previous values of $y_{t-1}, \ldots, y_{t-p}$ multiplied by their respective coefficients $\phi_1, \ldots, \phi p$, ε_t is the "innovation" at time t that encompasses everything not explained by the rest of the model, and c is a constant. Cryer and Chan (2008, sec. 4.3, pg. 70) show that an Auto-Regressive (AR) process can be represented as a general linear process. The *lag* operator L, $L^i \varepsilon_t = \varepsilon_{t-i}$, is also known as the *backshift* operator. As the name implies, an AR process is a *regression of a variable on itself.*

We will restrict the discussion to the *stationary* AR process, where the roots of the characteristic AR equation

$$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0, \tag{A.7}$$

are outside the unit disk in the complex plane.

A.3.2 Moving Average (MA) Models

In models where only past q white noise terms are considered, we get a Moving Average (MA) process

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{A.8}$$

or, in lag operator polynomial notation,

$$y(t) = \mu + \theta(L) \varepsilon_t, \tag{A.9}$$

where, $\theta(L)$ is the lag operator polynomial representing the linear combination of past white noise terms $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$ multiplied by their respective coefficients $\theta_1,\ldots,\theta_q$, ε_t is a series of past 'innovations' or noise, and $\mu(=0)$ is the unconditional mean of y. Another way of stating this is that only a finite number q of weights ψ_i, from Equation (A.2), are non-zero. We will restrict the discussion to *invertible* Moving Average (MA) processes, where the roots of the characteristic MA equation

$$1-\theta_1 x+\theta_2 x^2+\cdots+\theta_q x^q=0, \tag{A.10}$$

are outside the unit circle in the complex plane. This gives the useful result that there is only one set of coefficients that "yield an invertible MA process with a given autocorrelation function" (Cryer and Chan 2008).

A.3.3 Auto-Regressive Moving Average (ARMA) Models

In an $ARMA\ (p,q)$ model, the value of a time series at a certain point in time is predicted by a polynomial composed of two parts

$$y_t=c+\left[\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}\right]+\left[\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q}\right], \tag{A.11}$$

or, in lag operator polynomial notation,

$$\phi(L)y(t)=c+\theta(L)\varepsilon_t. \tag{A.12}$$

The same stationarity and invertibility conditions apply to an ARMA model as to its constituent AR and MA models. We are only concerned in this work with causal and invertible models, i.e., ones that predict a present value based only on past values. Since differencing is not used in this thesis, we do not consider Auto-Regressive Integrated Moving Average (ARIMA) models.

A.3.4 Homoscedasticity and Conditional Variance Models

Keeping in mind that we are only considering *conditional mean* models, it is essential that the time series be homoscedastic. If the variance of the residuals changes, the residuals are not homoscedastic (they are instead heteroscedastic). For a time series y_t, Cryer and Chan (ibidem) define the conditional variance of y_t given past values $y_{t-1}, y_{t-2}, \ldots$ as the variance of the deviation of y_t from its conditional mean $E(y_t|y_{t-1}, y_{t-2}, \ldots)$. Models for heteroscedastic time series exist, of which common types are the Auto-Regressive Conditional Heteroscedasticity (ARCH) and Generalised Auto-Regressive Conditional Heteroscedasticity (GARCH) models, proposed by Engle (1982), and Bollerslev (1986) and Taylor (1986), respectively (Box, Jenkins et al. 2008; Cryer and Chan 2008; Shumway and Stoffer 2011).

A.3.5 Script Snippet for Solar Series Bootstrap

```
% Start by taking the daily mean of the raw synthetic tdb values.
tdbsyn.daily.syn.means = mean(tdbsyn.daily.syn.raw);

% Reshape the synthetic and tmy months to separate the means by month
SynMonths_res = reshape(SynMonths,24,[]);
SynMonths_res = (SynMonths_res(1,:))';
TMYmonths_res = reshape(tmytable.Month,24,[]);
TMYmonths_res = (TMYmonths_res(1,:))';

% It is necessary to do a nearest neighbour bootstrap month-by-month since it is
% possible that, for a given synthetic day, the closest daily mean in the TMY
% is in a different month. This means that the sunshine hours will be considerably
% different between the synthetic day and the selected ‘neighbouring’ TMY day.

% Find the 'k' nearest neighbours in the TMY daily means for each synthetic daily
  mean.
knn = 10;
nIDX = NaN(size(SynMonths_res));

% Cycle through every month

for m = 1:length(mt)

% Get the synthetic means for this month
synmeans = tdbsyn.daily.syn.means(SynMonths_res==m)';

% Get the tmy means for this month
tmymeans = dStats.tmy.tdb.mean(TMYmonths_res==m);

% Get the ending day of the current month - this is to make sure that when each
% month is sampled, the indices stored are for the YEAR, not just the month.

if m == 1
% January starts from idx = 1, so IDXadd = 0.
IDXadd = 0;
else
% Get the end-of-month days for all months preceding the current month.
CurrEOMdays =  eomday(year(mt(1:m-1)),month(mt(1:m-1)));
% Sum them to get the number of days that have passed (sum end-of-month days).
IDXadd = sum(CurrEOMdays);
end

```

```
% Get the knn nearest neighbours in the TMY record for each daily mean
% in the synthetic time series.
[temp1, ~] = knnsearch(tmymeans,synmeans, 'K', knn, 'Distance','mahalanobis');
% The mahalanobis distance is used since it is reasonably robust
% against the shape of the underlying distribution.

% Pick one nearest neighbour at random
nIDX(SynMonths_res==m) = temp1(:,randi([1 10]));

% Add the preceding days of the year
nIDX(SynMonths_res==m) = nIDX(SynMonths_res==m) + IDXadd;

end

clear m tmymeans synmeans temp1 temp2

% nIDX is the index of a day in the TMY record whose daily mean
% matches some day in the Synthetic series.

% Pick the corresponding daily sums
ghisyn.daily.sums = dStats.tmy.ghi.sum(nIDX);
% Record the daily means as well
ghisyn.daily.means = dStats.tmy.ghi.mean(nIDX);
% Reshape the hourly ghi to be separated by days
ghi_res = reshape(tmytable.GHI,24,[]);
% Pick the days corresponding to each nearest neighbour
ghisyn.daily.raw = ghi_res(:,nIDX);
```

A.4 Additional Results

Additional results are presented here for the three cities used as examples in this thesis – Geneva, New York, and Delhi (details in Table A.1). There are three stations in New York, and their results are compared side-by-side. As we discussed in Chapter 3, the quality of the weather file influences the quality of the generated data. In the case of Delhi, for example, the low values of Relative Humidity (RH) per month are not as low as the recorded values. The thresholds in the spell plots (Figures A.7 and 3.23) correspond to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) design temperatures. Temperatures corresponding to the 99.6, 99, 98, 2.0, 1.0, and 0.4 percentiles are presented here.

Table A.1 – *Details of the climates/stations described in the main chapter.*

City Name	Geneva	New York	New York	Delhi
Station	Cointrin	John F Kennedy	LaGuardia	Indira Gandhi
Code	GEN	NYC_JFK	NYC_LAG	DEL
Longitude	6.10	-73.80	-73.88	77.10
Latitude	46.23	40.65	40.78	28.57
ASHRAE	4B	5A	5A	1B
Koeppen-Geiger	Cfb/Dfb	Dfa/Dfb	Dfa/Dfb	Cwa

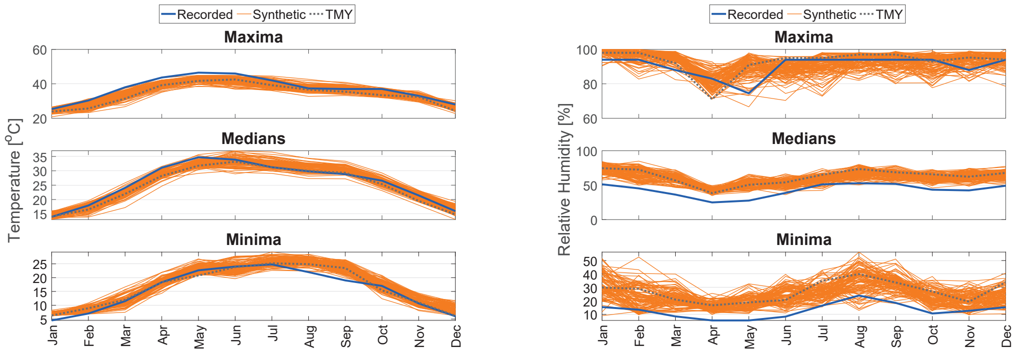


Figure A.3 – *TDB [top] and RH [bottom] ranges for Delhi. Unusually for the results presented, recorded highs in spring are* higher *than the synthetic ones.*

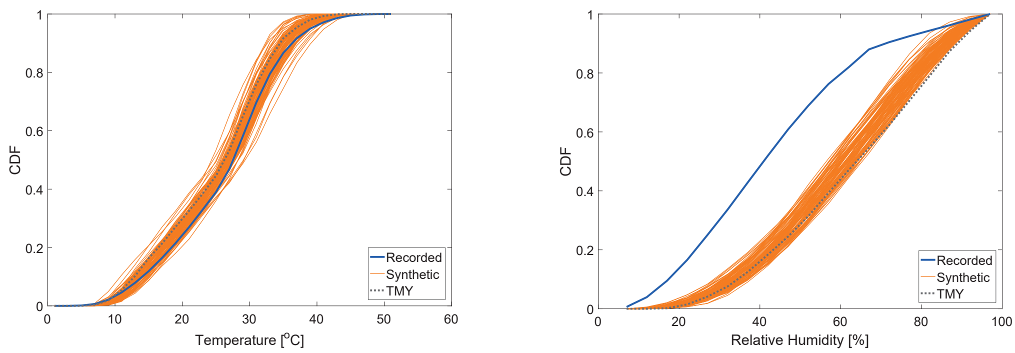


Figure A.4 – *TDB [top] and RH [bottom] empirical Cumulative Density Functions (eCDFs) for Delhi.*

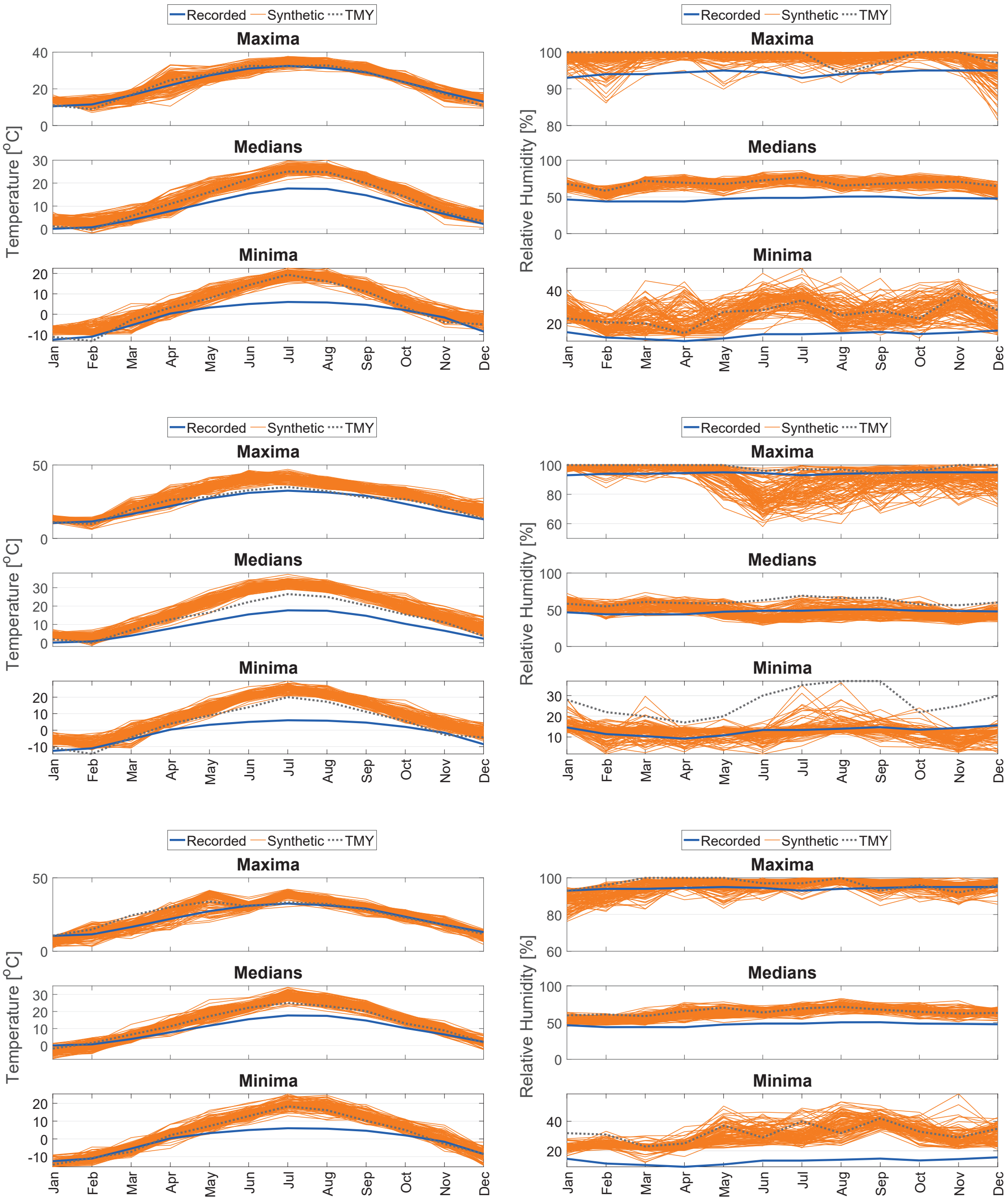
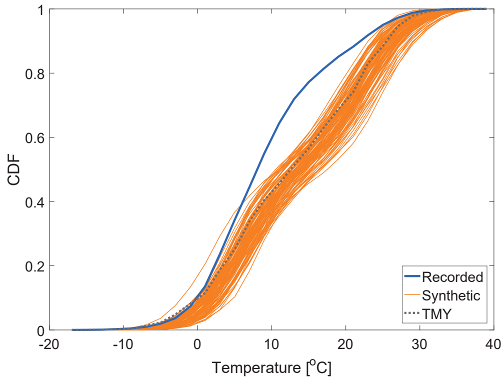
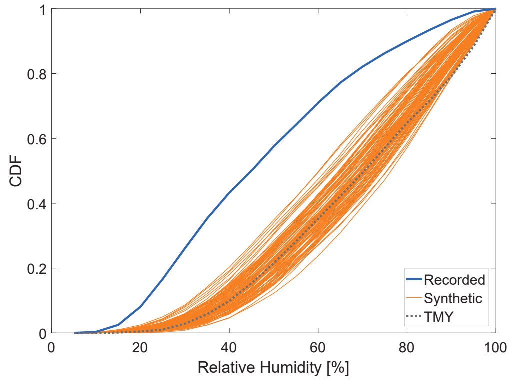


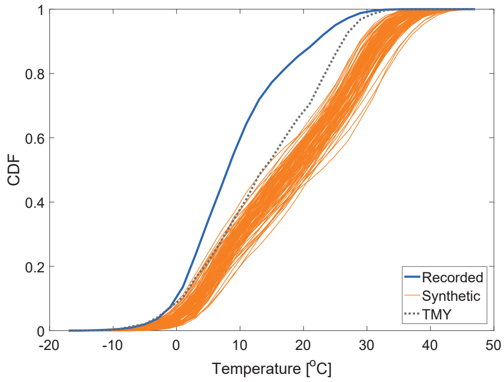
Figure A.5 – *Dry Bulb Temperature (TDB) [left] and Relative Humidity (RH) [right] extents for New York JFK [top], LAG [middle], and CPR [bottom].*



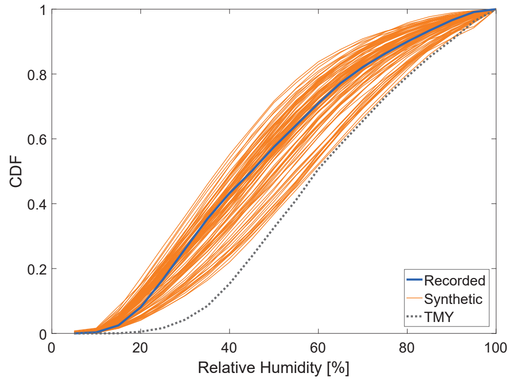
(a) *JFK, TDB*



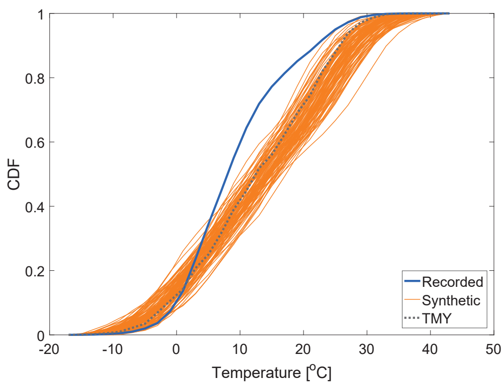
(b) *JFK, RH*



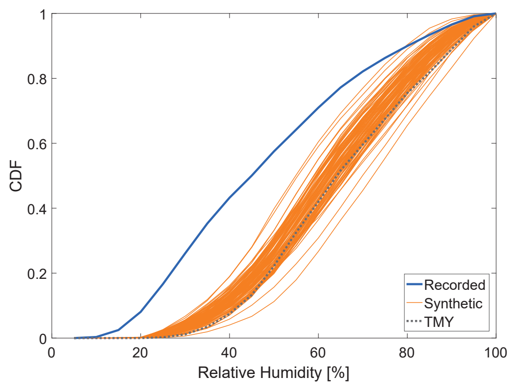
(c) *LAG, TDB*



(d) *LAG, RH*



(e) *CPR, TDB*



(f) *CPR, RH*

Figure A.6 – *TDB [left] and RH [right] eCDFs for New York.*

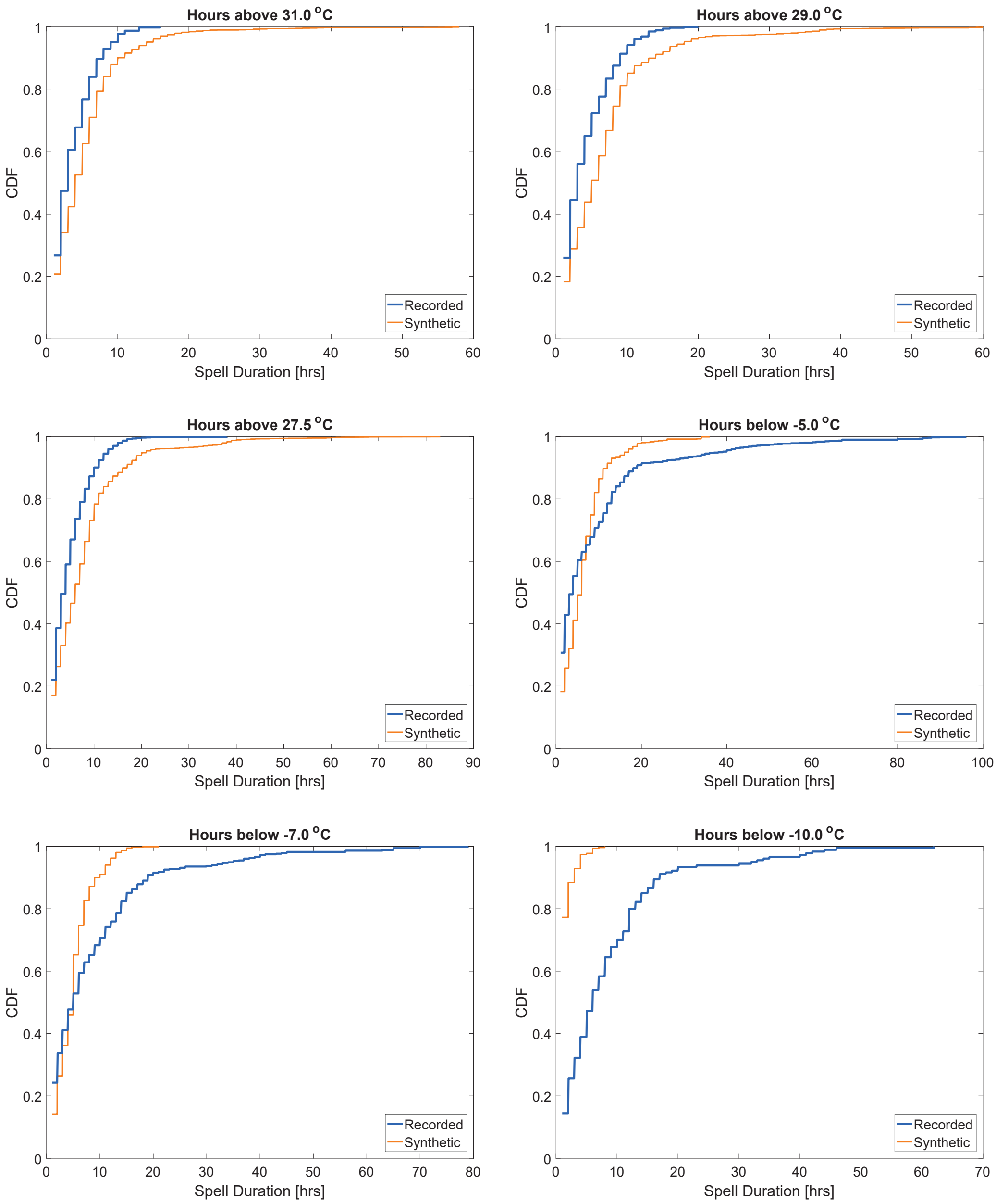


Figure A.7 – *eCDFs of the spell durations in recorded and synthetic data for New York JFK: [from top left] 99.6, 99, 98, 2.0, 1.0, and 0.4 percentiles.*

A.5 Weather Data Sources

Recorded/historical weather data was downloaded from multiple national and international agencies. An example MATLAB®script for reading the raw data file from MeteoSwiss, the Swiss meteorological service, is included below. In many cases, the data overlapped. For example, temperature data for Swiss locations was available with both the NCDC and MeteoSwiss. In these cases, the average of the two sources was used. The relevant MATLAB and R code is available with the archive copy of this thesis on the EPFL repository (infoscience.epfl.ch). The link to a GIT repository will also be found there.

- (Swiss) Federal Office of Meteorological and Climatology (MeteoSwiss 2014) – all data for locations in Switzerland.
- [United States] National Solar Radiation Database (NSRDB) (Wilcox 2012) – solar and temperature data for the United States.
- NREL India Solar Resource Data (Sengupta 2014) – solar data for India.
- Canadian Meteorological Service (Meteorological Service of Canada and National Research Council of Canada 2008) – solar and temperature data for Canada.
- NASA Remote Sensing Validation Data (al-Abbadi 2002) – solar, temperature, humidity data for Saudi Arabia.
- NCDC (NCDC/NOAA 2014) – temperature, humidity, and atmospheric pressure for all locations.

Typical weather year data was obtained from three sources:

- EnergyPlus Weather Data (NREL and USDOE 2015)[8];
- METEONORM (MN) (Remund, Mueller et al. 2012b);
- and, TMY3 (Wilcox and Marion 2008).

[8] The website has moved since we last accessed the data in 2012. This reference is to the new link. EnergyPlus format data can also be found at http://climate.onebuilding.org.

B Regression: Additional Concepts and Details

For results from other climates and buildings, besides the three presented here, as well as original code, the reader is directed to the supplementary documents provided with the archival copy of this thesis on the EPFL repository (infoscience.epfl.ch). The text of this part of the thesis builds upon extensive discussions with Emilie Nault, and the work previously published in Nault (2016) and Nault, Rastogi et al. (2015). Figure 4.2 (Chapter 4, case 1) is adapted from Chinazzo (2014). The original energy model for the home (case 1) was made with Soenke F. Horn, and first published in Rastogi, Horn et al. (2013).

B.1 Confidence Intervals

Since the ideas of confidence intervals, prediction intervals, and variability intervals are crucial to the understanding and use of uncertainty and sensitivity analysis, a brief explanation of their use and interpretation is called for. In this section, we give a general conceptual explanation of these intervals, including details of the prediction intervals plotted in the regression plots (e.g., Figure 4.7) when appropriate. In the context of regression models (Chapter 4), we use prediction intervals, and all of these intervals are constructed for specific confidence levels, e.g., $\alpha = 0.5$ gives 95% confidence intervals or prediction intervals.

Say we fit a curve to data as follows

$$y = f(\boldsymbol{x}) + \varepsilon, \tag{B.1}$$

where y is the dependent variable, $\boldsymbol{x}$ is a vector of predictors (independent variables, regressors), $f(\boldsymbol{x})$ is the unknown underlying function from which the data supposedly arises (and which a fitted curve should approximate), and ε is random error. Since the true underlying curve is unknown in our application, Equation (B.1) is replaced with that of a fitted curve

$$\hat{y} = \hat{f}(\boldsymbol{x}) + \hat{\varepsilon}, \tag{B.2}$$

The model is queried at an unknown point

$$\hat{y}_{n+1} = \hat{f}(\boldsymbol{x}_{n+1}) + \hat{\varepsilon}_{n+1}, \tag{B.3}$$

where, the model was fit using n training points (i.e., $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$), $\boldsymbol{x}_{n+1}$ is the vector of predictors at a new query point, and $\hat{\varepsilon}_{n+1}$ is the associated error (which is independent of errors at all other points, $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$). It is possible to construct prediction intervals on both the fitted function/curve, $\hat{f}(\cdot)$, and new predicted outputs, $\hat{y}$. MATLAB® lists these two options to determine whether the prediction intervals should be for the fitted function (`curve` or fitted mean values) or the predicted value (`observations`). 'Observation' prediction intervals, which we use in this thesis for linear model plots, give wider bounds because "the error in a new observation is equal to the error in the estimated mean value, plus the variability in the observation from the true mean" (The MathWorks, Inc. 2015). In the formulations, the difference between the two types of confidence intervals is the presence of an extra term in the confidence intervals for observations.

A second option is for whether the prediction intervals should hold true for *all* predictor values "simultaneously", or not. The simultaneous option, which is what is used in this thesis, gives wider bounds, because it is a more stringent requirement that the specific bounds enclose the "entire curve" or all predicted values rather than just at a single predictor value (ibidem). In this case, the prediction is

$$\hat{y}_{n+1} = \hat{f}(x_{n+1}) + \hat{\varepsilon}, \tag{B.4}$$

where $\hat{\varepsilon}$ is now the error over all the values (contrasted with the error for a specific set of predictors in Equation (B.2)). In Gaussian linear models, the critical statistic for the simultaneous values comes from an inverse F cumulative distribution, while that for the non-simultaneous comes from an inverse of Student's t cumulative distribution (ibidem).

In the context of the classical regression examples discussed in this thesis, the plotted intervals imply that at any given 'regression-predicted' value of the output ($\hat{y}$), the given bounds contain the 'true' value (y) in about 0.95 or 95% of experiments or model fits. So, we are plotting *prediction intervals*. If the bounds are far apart, then the model prediction is less useful even if it has, technically, the same accuracy. For the Gaussian Process regression models, the fixed bounds of the prediction intervals contain a Gaussian random variable (with a mean and variance determined by the Gaussian Process). That is, the probability that y lies within the given bounds is 95%, with a Gaussian distribution.

B.2 More Gaussian Processes

B.2.1 GP Kernels: Additional Details

A popular kernel is the squared exponential function (SqE) function

$$k_{se}(x_i, x_j) = \sigma_f^2 \exp\left[\frac{-(x_i - x_j)^2}{2\sigma_l^2}\right], \tag{B.5}$$

where x_i and x_j are two sample points, σ_f is roughly representative of the intrinsic variance of the response variable, and σ_l is the *characteristic length scale*. The characteristic length scale is broadly representative of the roughness in the latent function due to each predictor. Drawing a parallel with Principal Component Analysis (PCA), we could say that the characteristic length scale is broadly indicative of the amount of variance that would be explained by a particular predictor. The maximum covariance is σ_f^2, since as x_i approaches x_j, or $x_i - x_j \approx 0$, then $k(x_i, x_j) \approx \sigma_f^2$. A high value for σ_f^2 implies that the output shows more variability.

The SqE, also called the radial basis function (RBF) is popular because it is infinitely differentiable, flexibly integrable, and broadly applicable (Duvenaud 2014). A natural extension of the squared exponential function is to use different characteristic length scales for each independent variable – the automatic relevance determination (ARD) squared exponential function,

$$k_{se,ard}(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{m=1}^{d} \frac{(x_{i,m} - x_{j,m})^2}{\sigma_m^2}\right), \tag{B.6}$$

where σ_m is an element of a $d \times 1$ vector of characteristic length scales, one for each independent variable. Given our discussion so far about the heterogeneous responses of building simulation outputs to different variables, the ARD function is attractive. However, we will see later that it does not, in fact, increase the predictive power sufficiently to merit the additional $d-1$ parameters. The ARD SqE kernel can be thought of as a product of several SqE kernels. Using the conditions discussed in Duvenaud (2014), we examine the suitability of the squared exponential functions as covariance functions. The functions are weakly stationary, or "invariant to translations in the input space", since they are functions of only the distance between two query points, $h = x_i - x_j$. Similarly, they are *isotropic* because of the square around the $(x_i - x_j)$ term (Rasmussen and Williams 2006, chpt. 4). There are several other functions that meet these requirements, including the *Matern kernels*

$$k_{m32}(x_i, x_j) = \sigma^2\left(1 + \frac{\sqrt{3}r}{\sigma_l}\right)\exp\left(-\frac{\sqrt{3}r}{\sigma_l}\right), \tag{B.7}$$

$$k_{m52}(x_i, x_j) = \sigma^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{\sqrt{5}r^2}{\sigma_l^2}\right)\exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right), \tag{B.8}$$

which may be called the $\frac{3}{2}$ and $\frac{5}{2}$ kernels respectively, and where $r = |x_i - x_j|$. Kernels can be periodic,

$$k_{per}(x_i, x_j) = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|x_i - x_j|/p)}{\sigma_l^2}\right), \tag{B.9}$$

where p is the period and σ_l is the same length scale parameter as before; or even linear

$$k_{lin}(x_i, x_j) = \sigma_b^2 + \sigma_v^2(x_i - c)(x_j - c), \tag{B.10}$$

where c is the 'horizontal offset' of the posterior distribution of the response surface and σ_b^2 is the 'vertical offset' of the response surface. The linear kernel is different for one important reason: it is non-stationary. Recall that a "stationary covariance function is one that only depends on the relative position of its two inputs, and not on their absolute location" (Duvenaud 2014). Each kernel has an ARD version. Finally, kernels may even be combined by multiplication and addition.

B.2.2 Implementation

The implementation of Gaussian Process regression used in this thesis is the function `fitrgp` and its associated routines in MATLAB®. We present some important aspects of the implementation for this thesis.

The function offers a choice of the squared exponential and Matern kernels, and their ARD versions, though a custom kernel is also permitted. Given that the focus of this thesis is the *application* of Gaussian Process regression, we will by and large stick to the inbuilt options. Sections B.2.1 and 4.4.1.3 discuss the differences between the different kernel choices. In this thesis we assume that each known data point (i.e., simulation result) is exact, though that is not necessary for Gaussian Process regression. In fact, assuming that each data point comes with some uncertainty could form the basis for alternative future implementations (Section 5.3).

B.2.2.1 Fit Method

The 'fit method' option governs how the hyper-parameters of a fit are selected. The software offers five choices for fit method: no estimation, using the input parameter values as the final values; exact fitting, the slowest option, using all predictors and data points; using a subset of the data points or regressands; using a subset of predictors/regressors; and, fully independent conditional approximation, which is an improvement on the subset of regressors approximation. Initially, we tried the exact method, which relies on Maximum Likelihood Estimation (MLE). That is, the function selects the optimal kernel parameters by maximizing a likelihood function given the input hyper-parameter vector $\boldsymbol{\theta}$, like in the classical regression techniques discussed previously (Section 4.3.1). For the optimisation routines, the starting value we usually assigned to σ_f was the standard deviation of the response variable. For the SqE kernel, the starting characteristic length scale was one of the elements of a $d \times 1$ vector of the standard deviations of the inputs. This choice severely over-fit the functions to the training data, and slowed down considerably when the training data set exceeded a few hundred points. The subset options sped up the calculations but the gain was not justified by the significant loss of fit quality, even compared to the over-fitting of the Maximum Likelihood Estimate (MLE)-based estimation. An interesting analogue to these issues of over-fitting and unrepresentative sampling is in the choice of the master training data set: using just typical weather data for training is severely insufficient (Sections 4.3.3.2 and 4.4.2).

Using k-fold cross-validation with the 'no estimate' option achieved the best predictive performance over the master testing set (see Section 4.4.1.2). This option makes the `fitrgp` function use the hyper-parameters input by the user.

B.2.2.2 Predict Method

This input has five options, more or less aligned with the fit method: exact, block coordinate descent, subset of data points, subset of predictors, and fully independent conditional approximation. We picked the 'exact' option, which is the most flexible but also the slowest of the methods.

B.2.2.3 Basis Function

This is an option for on-the-fly basis transformation. If a transformation is requested, the Gaussian Process (GP) is fitted to transformed data in a new basis space $\mathbb{R}^p$. The basis function can range from a constant to any custom function. In the software documentation, the response is described as a linear combination of the basis function and the latent function

$$h(x)^T \boldsymbol{\beta} + f(x), \tag{B.11}$$

where $\boldsymbol{\beta}$ is a vector of unknown coefficients and $f(x)$ is the zero-mean Gaussian Process. This equation is equivalent to the *universal Kriging model* (Christensen 1991). Through trial-and-error, we found that a *constant* basis transform was enough (subtracting $\boldsymbol{\beta}$), which is equivalent to the *simple Kriging model* (ibidem).

B.2.2.4 Standardization and Regularization

Together, these two options replicate the z-score step. 'Standardize' asks the software to centre and scale the data, and regularization divides by the given standard deviation. These are not strictly necessary, but we used the 'standardize' step anyway to maintain robustness to input errors (like not using the z-score of one input).

B.2.3 Results: SQE Kernel

In this section, we present some additional results from Gaussian Process regression using the SqE kernel. These results are from the second fitting routine, i.e., using only typical and recorded weather files for training,. Compare these results to those presented in Section 4.4.2.2 for the third routine, i.e., Best Fit, where the training weather files are selected randomly (e.g., Figures 4.13, 4.14, 4.17 and 4.18). We can see that the approach of using only typical and recorded weather data for training performs worse than using synthetic data as well. The residuals in the graphs in this section are higher (e.g., Figure B.5), and the predicted-simulated plots also look worse (e.g., Figure B.1). Note that the plots in this section are all presented in terms of z-scores, like Section 4.4.2.2, which is why the quantities are all dimensionless. See Section 4.4.2.1 for an explanation of the plots presented here.

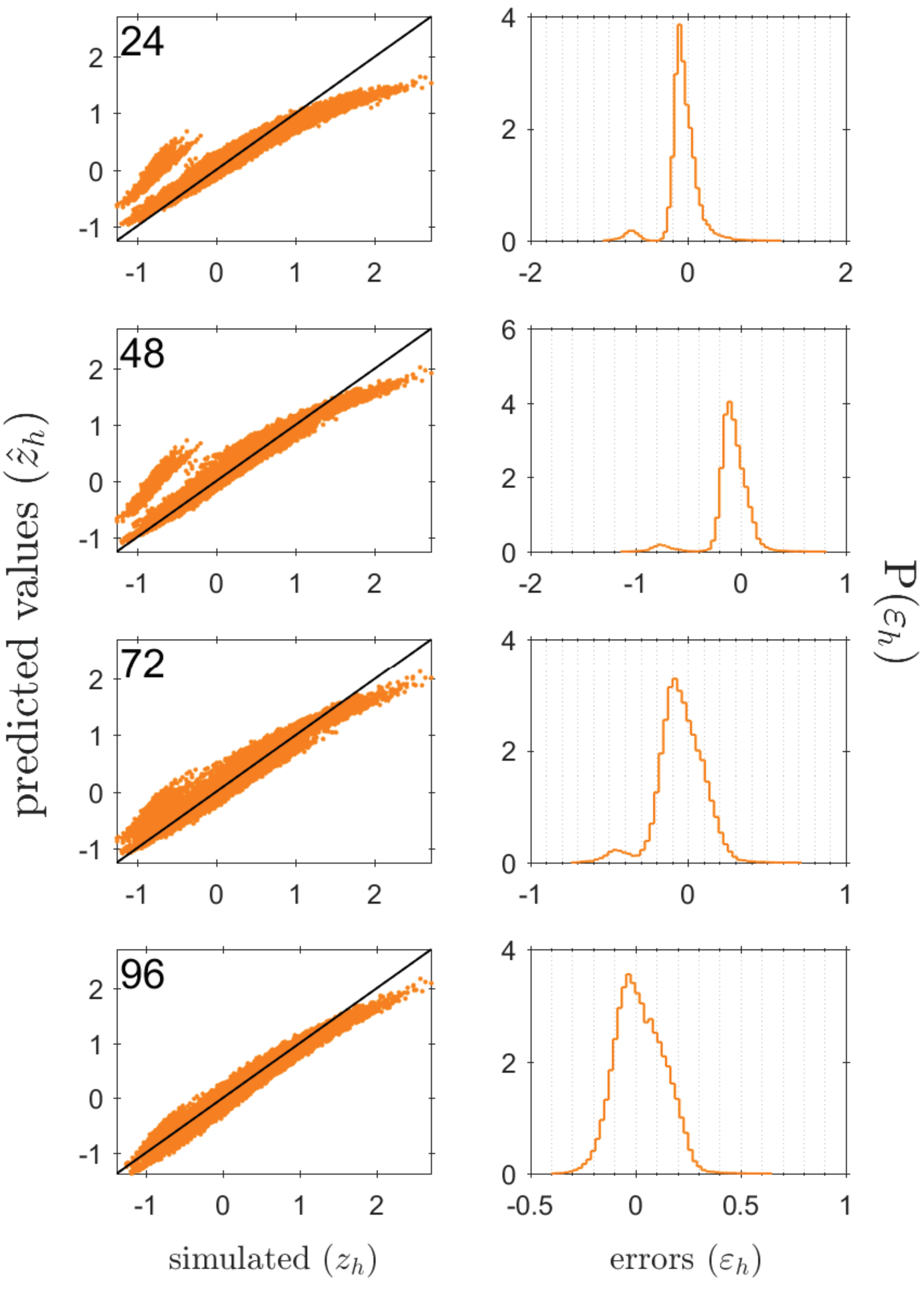


Figure B.1 – *Case 1, Second Fit – heating. Predictions [left] and residuals [right]. These are (unit-less) z-scores, not the original data.*

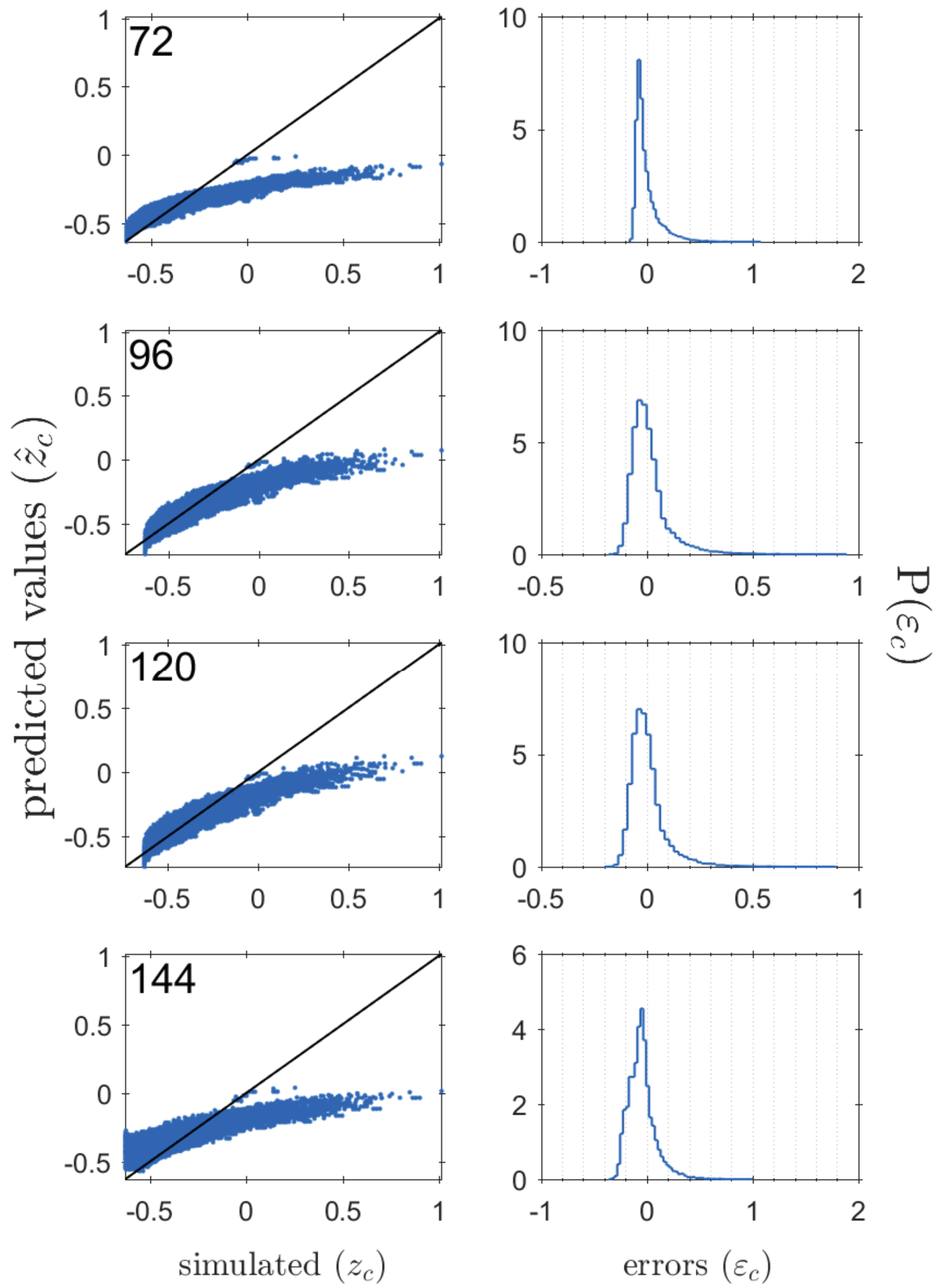


Figure B.2 – *Case 1, Second Fit – cooling. Predictions [left] and residuals [right]. These are z-scores, not the original data.*

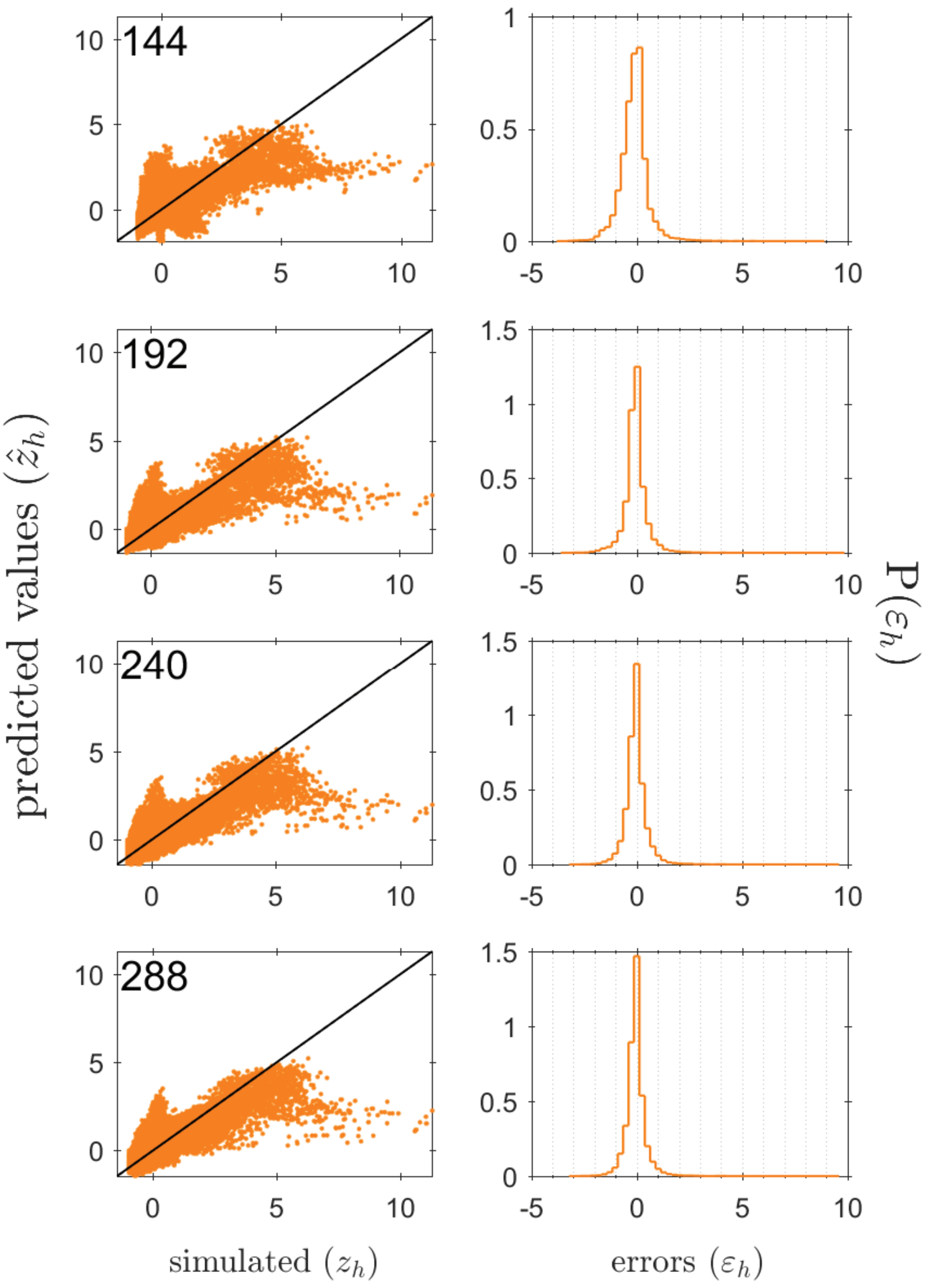


Figure B.3 – *Case 2, Second Fit – heating. Predictions [left] and residuals [right]. These are z-scores, not the original data.*

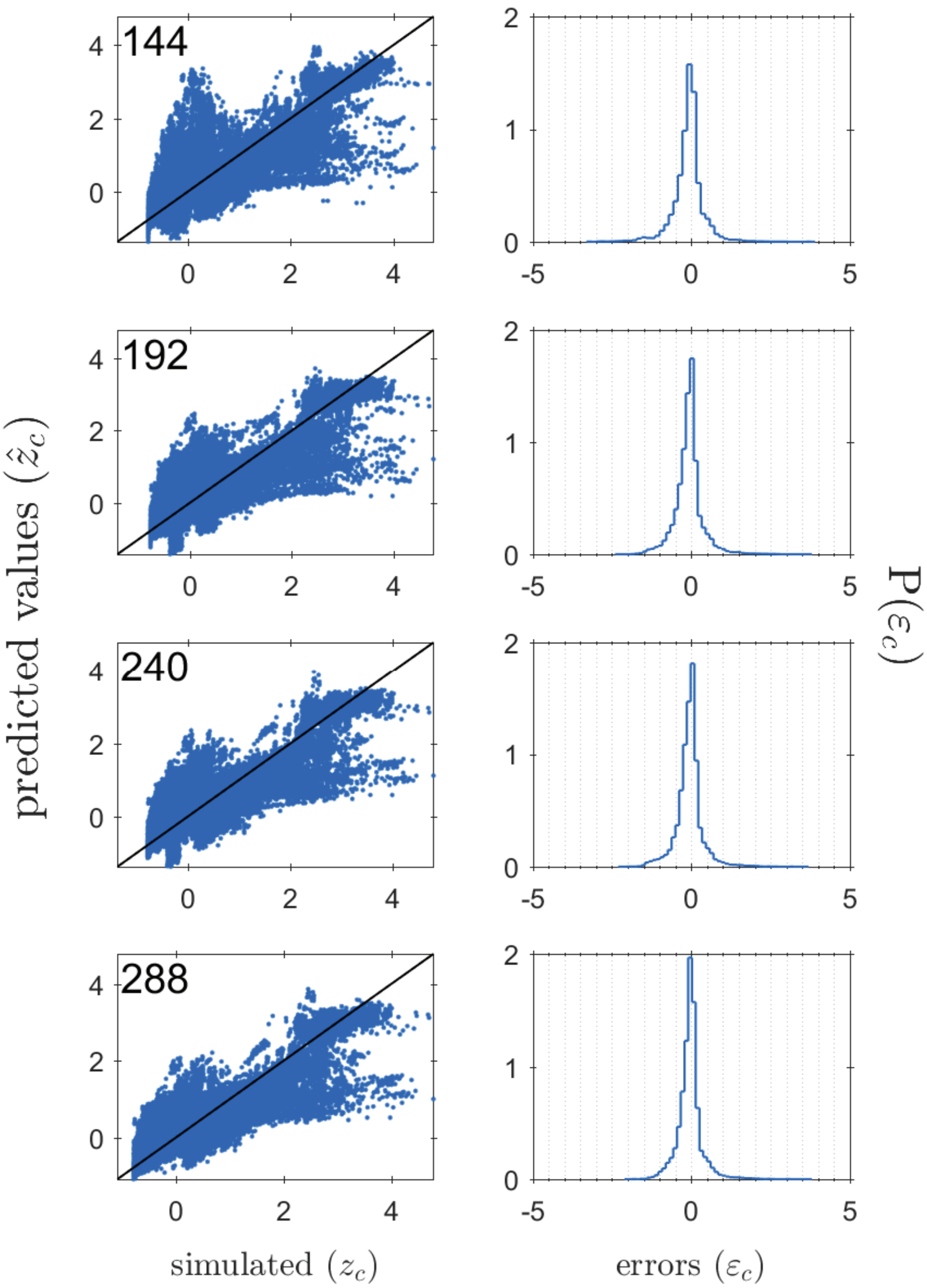


Figure B.4 – *Case 2, Second Fit – cooling. Predictions [left] and residuals [right]. These are z-scores, not the original data.*

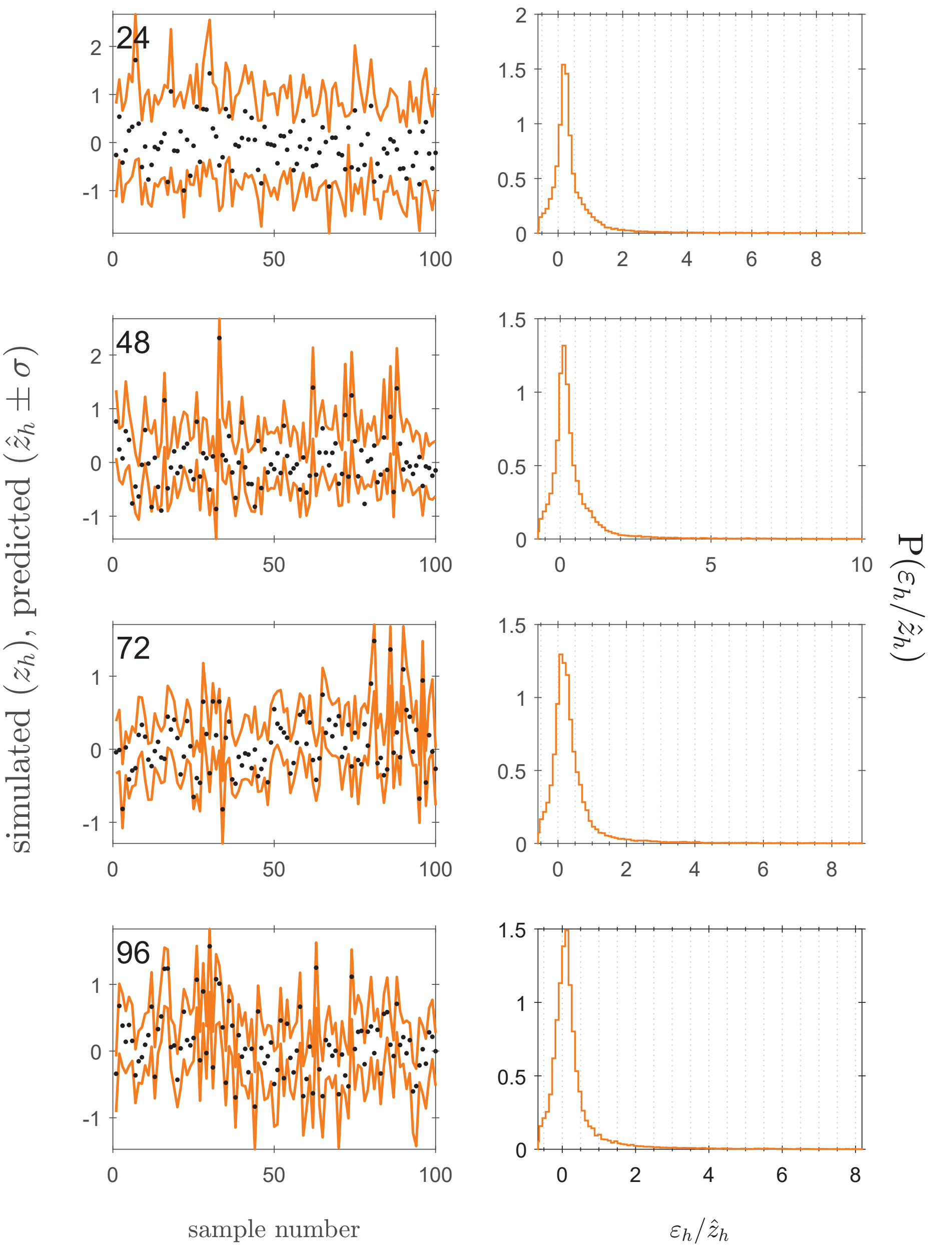


Figure B.5 – *Case 1, Second Fit – cooling. Prediction interval enclosing simulated values (black dots) [left]. Ratio of residuals to predictions [right]. These are z-scores, not the original data.*

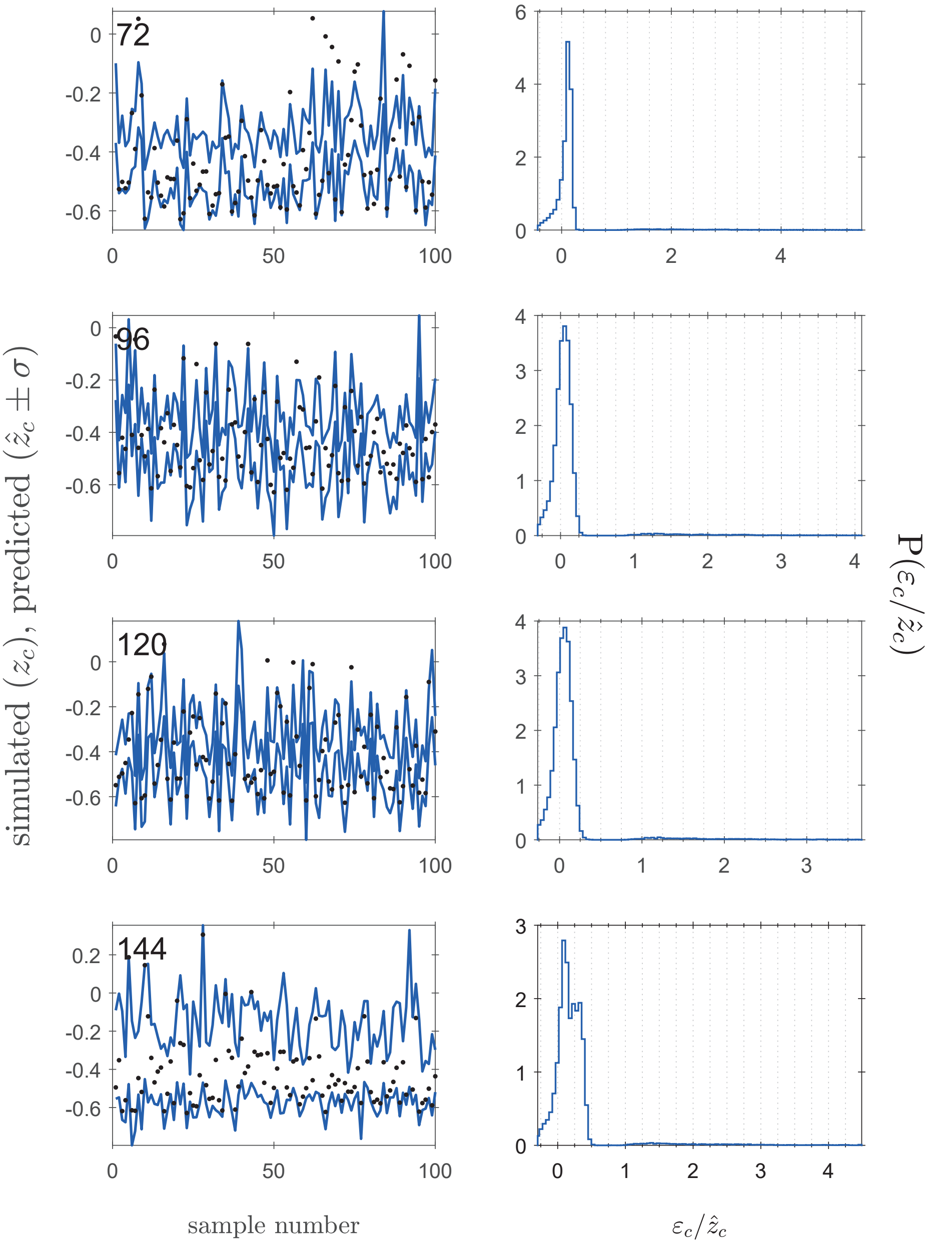


Figure B.6 – *Case 1, Second Fit – cooling. Prediction interval enclosing simulated values (black dots) [left]. Ratio of residuals to predictions [right]. These are z-scores, not the original data.*

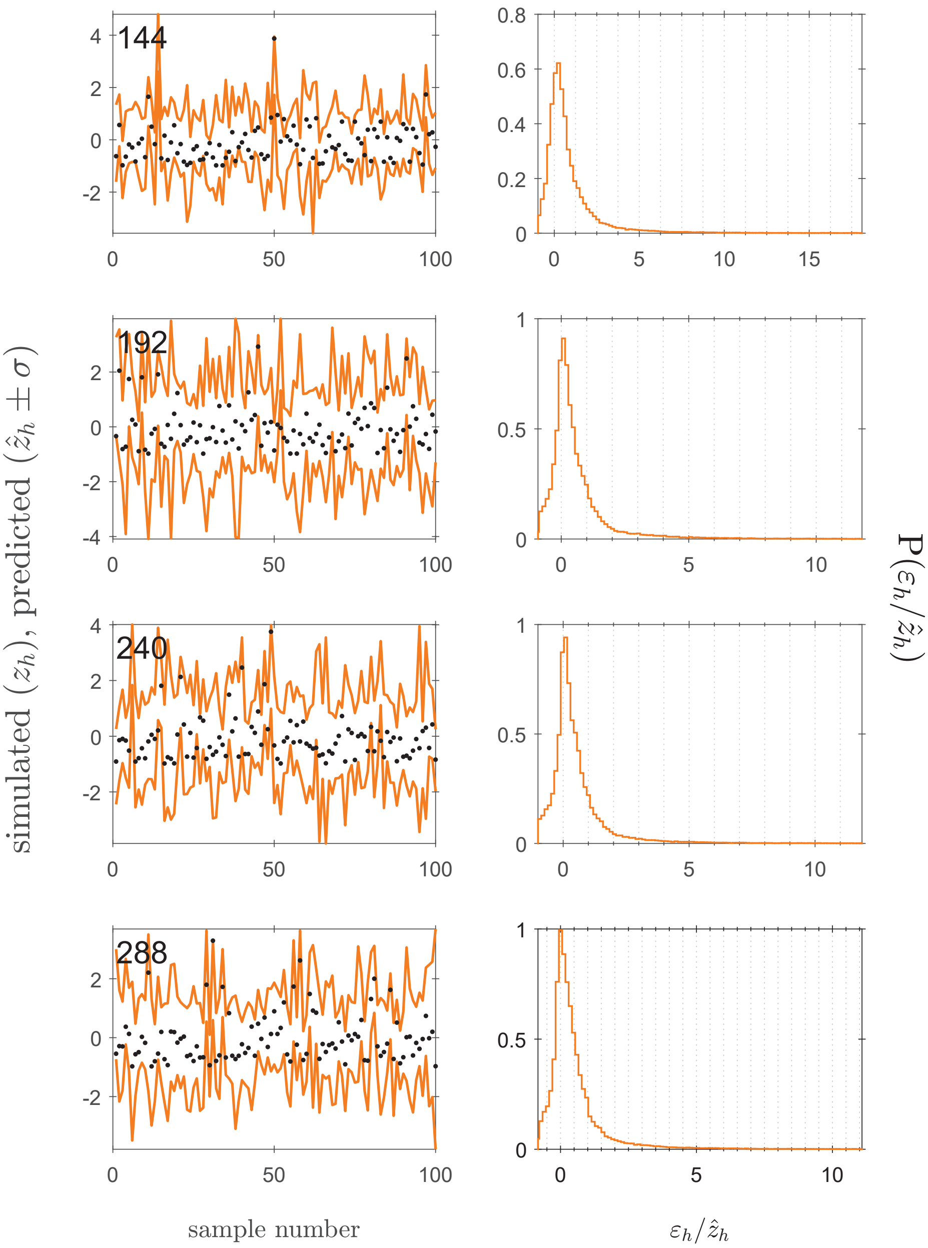


Figure B.7 – *Case 2, Second Fit – heating. Prediction interval enclosing simulated values (black dots) [left]. Ratio of residuals to predictions [right]. These are z-scores, not the original data.*

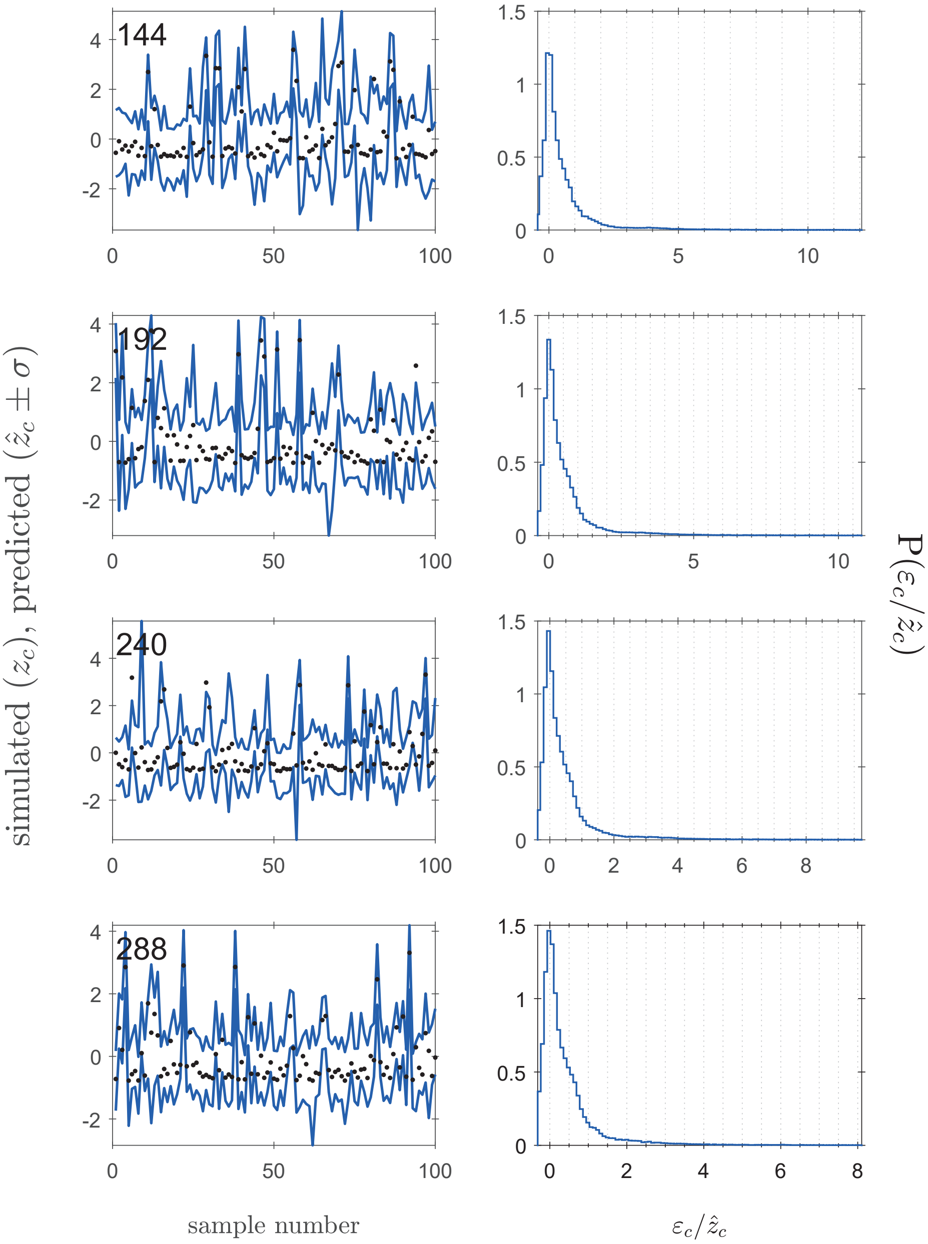


Figure B.8 – *Case 2, Second Fit – cooling. Prediction interval enclosing simulated values (black dots) [left]. Ratio of residuals to predictions [right]. These are z-scores, not the original data.*

B.2.4 Results: ARD Kernel

Compare the plots presented here to those presented in Sections B.2.3 and 4.4.2.2. The plots presented here are only for the third routine (Best Fit). The ARD kernel assigns a different length scale (σ_l, see Section 4.4.1) for each covariate or input/independent variable. The promise of this approach, in assigning differing covariance characteristics to each independent input variable, is not borne out in our example. The primary issue is over-fitting: the Gaussian Process regression model is too finely tuned to the training data. The predicted-simulated plots either show a tendency to predict the mean (e.g., Figure B.9) or a large spread for the same value on the x-axis (e.g., Figure B.10).

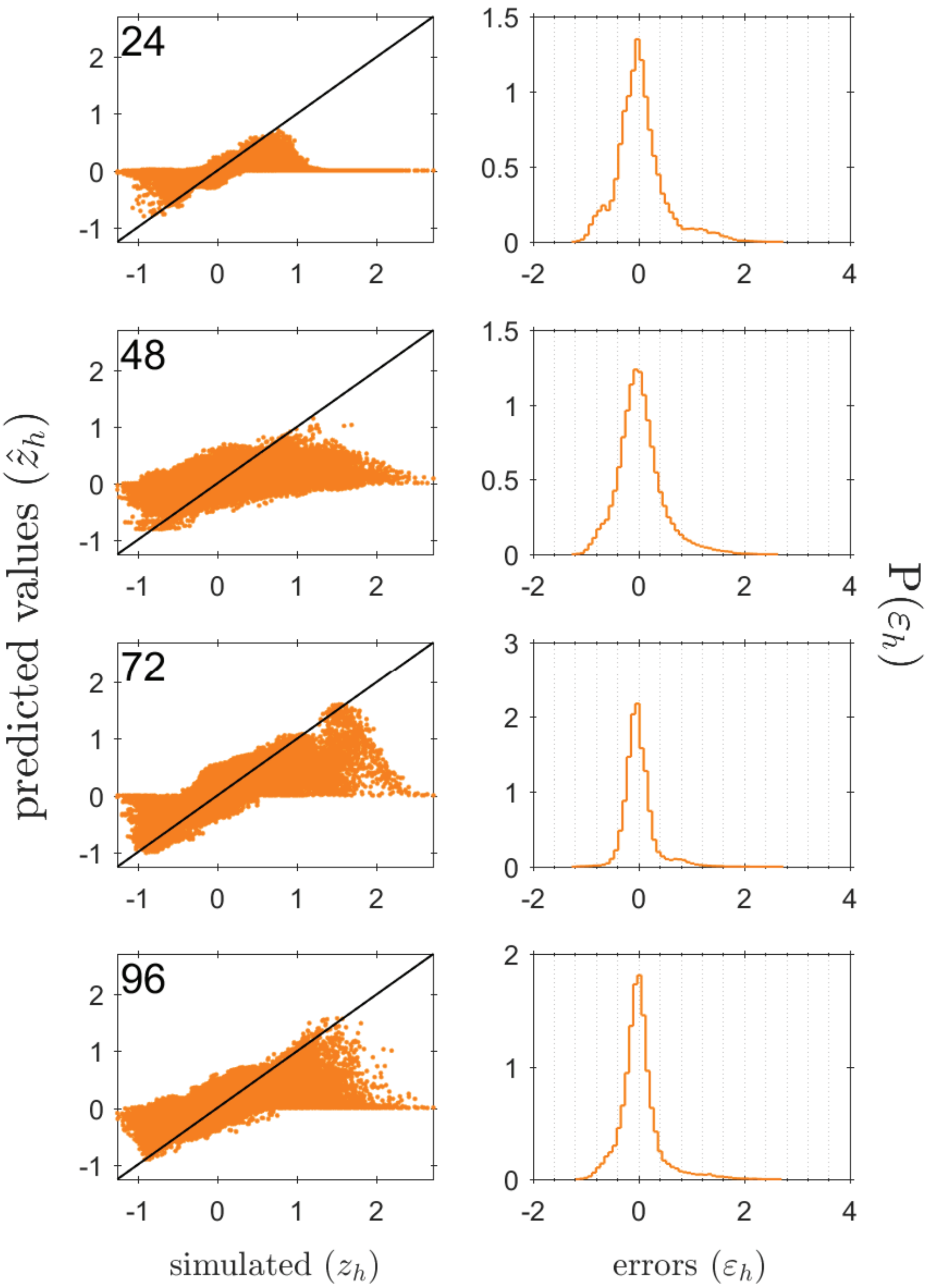


Figure B.9 – *Case 1, Best Fit, ARD kernel – heating. Neither model performs acceptably. These are z-scores, not the original data.*

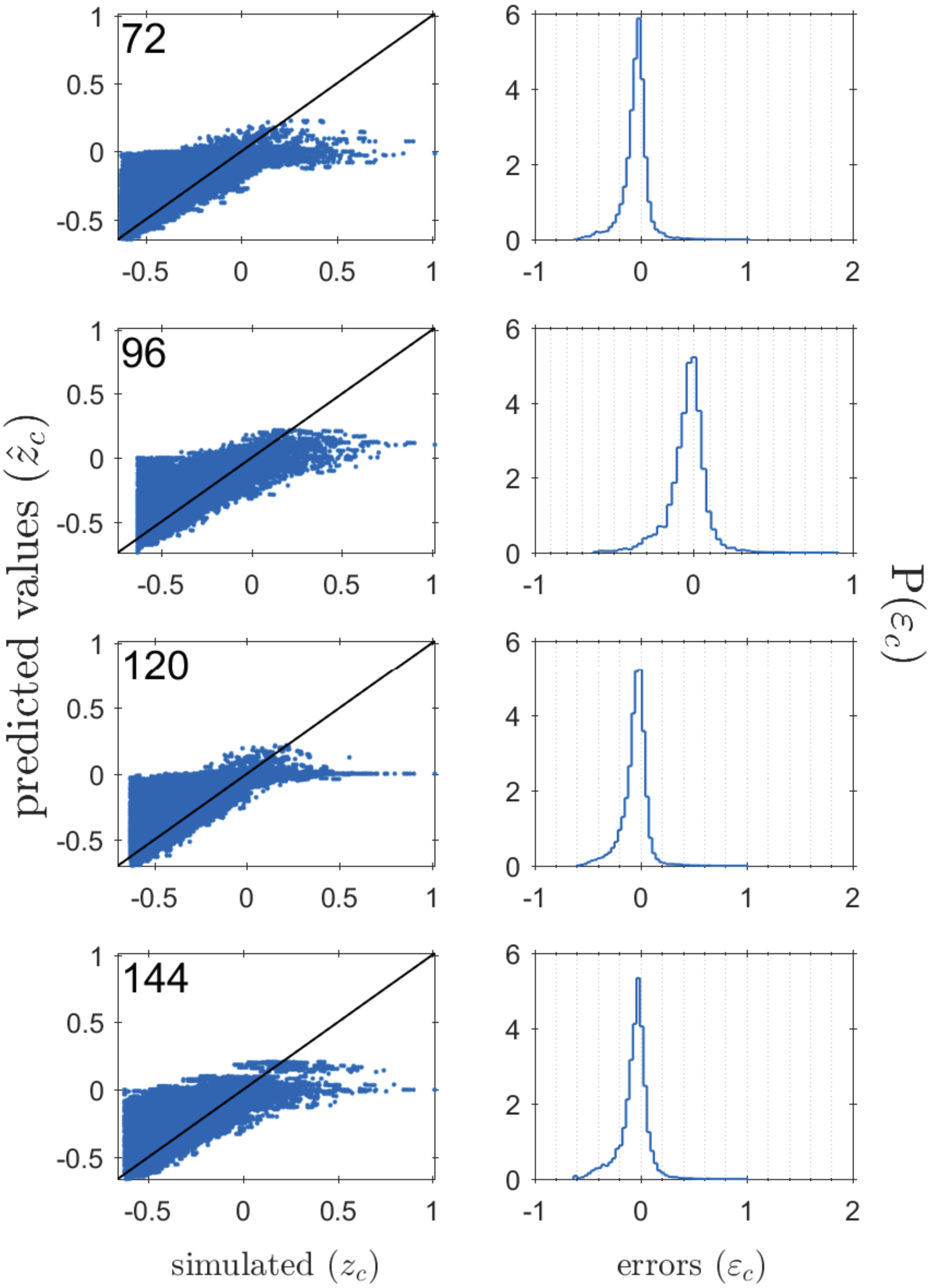


Figure B.10 – *Case 1, Best Fit, ARD kernel – cooling. Neither model performs acceptably. These are z-scores, not the original data.*

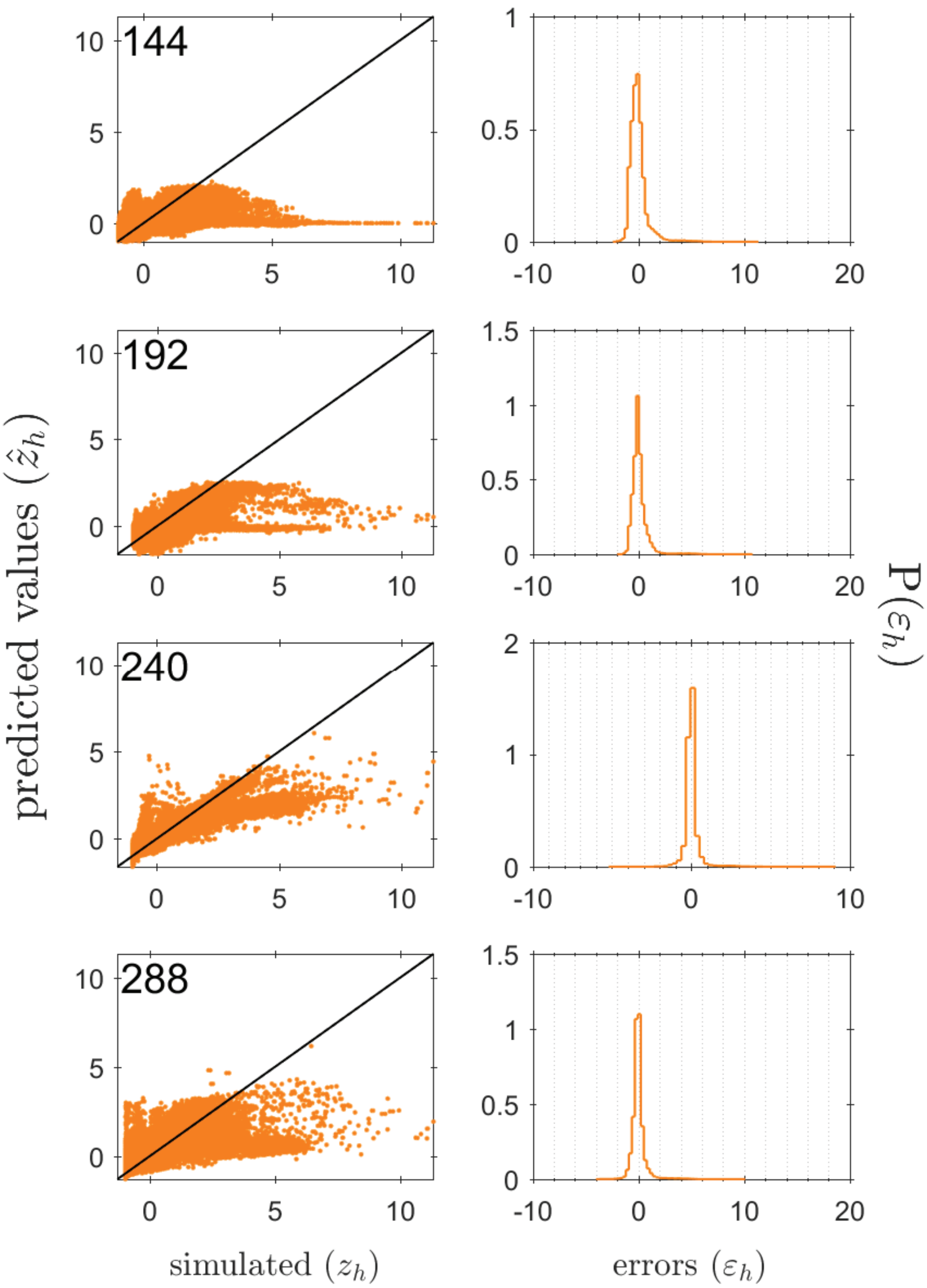


Figure B.11 – *Case 2, Best Fit, ARD kernel – heating. These are z-scores, not the original data.*

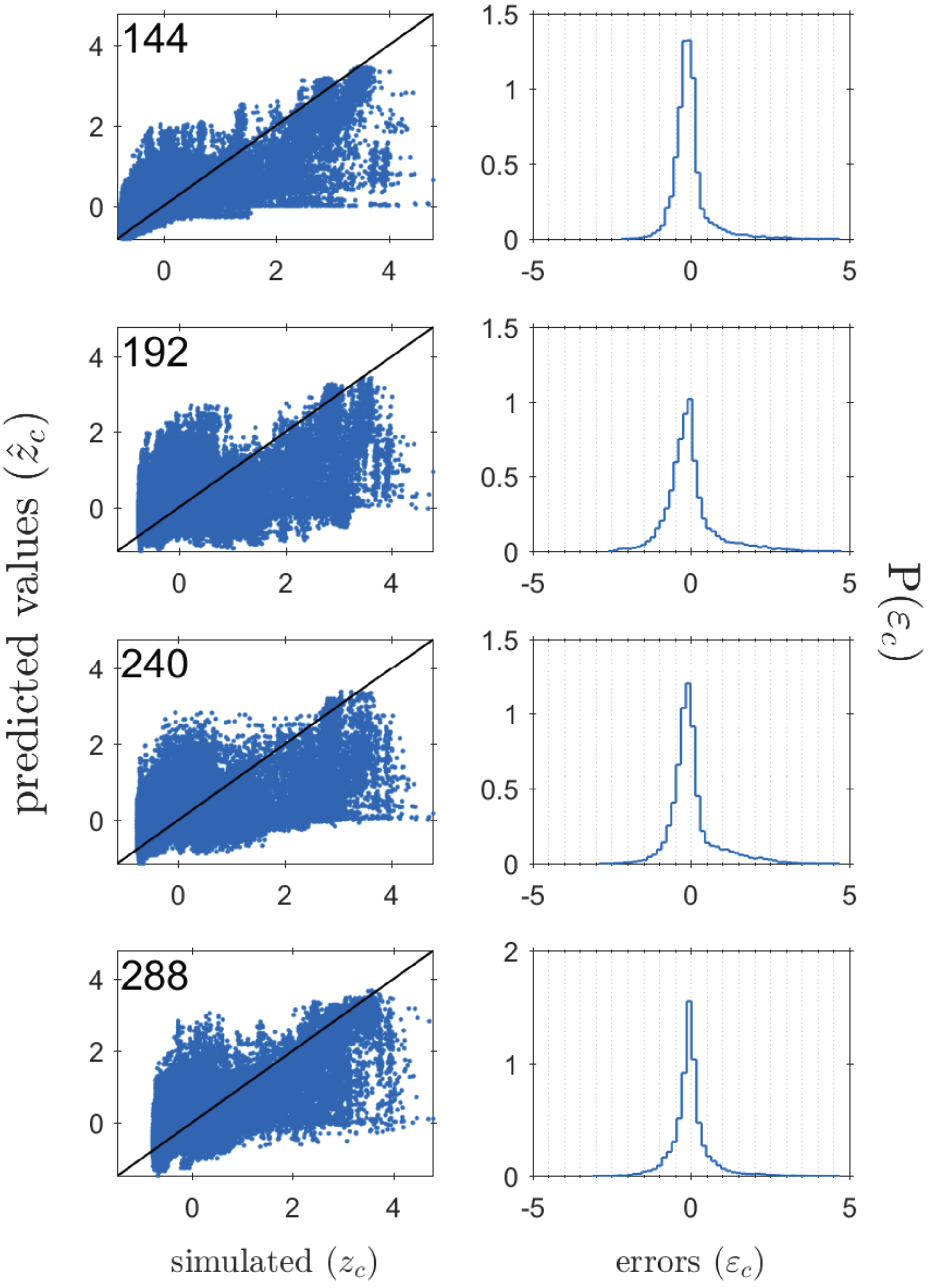


Figure B.12 – *Case 2, Best Fit, ARD kernel – cooling. These are z-scores, not the original data.*

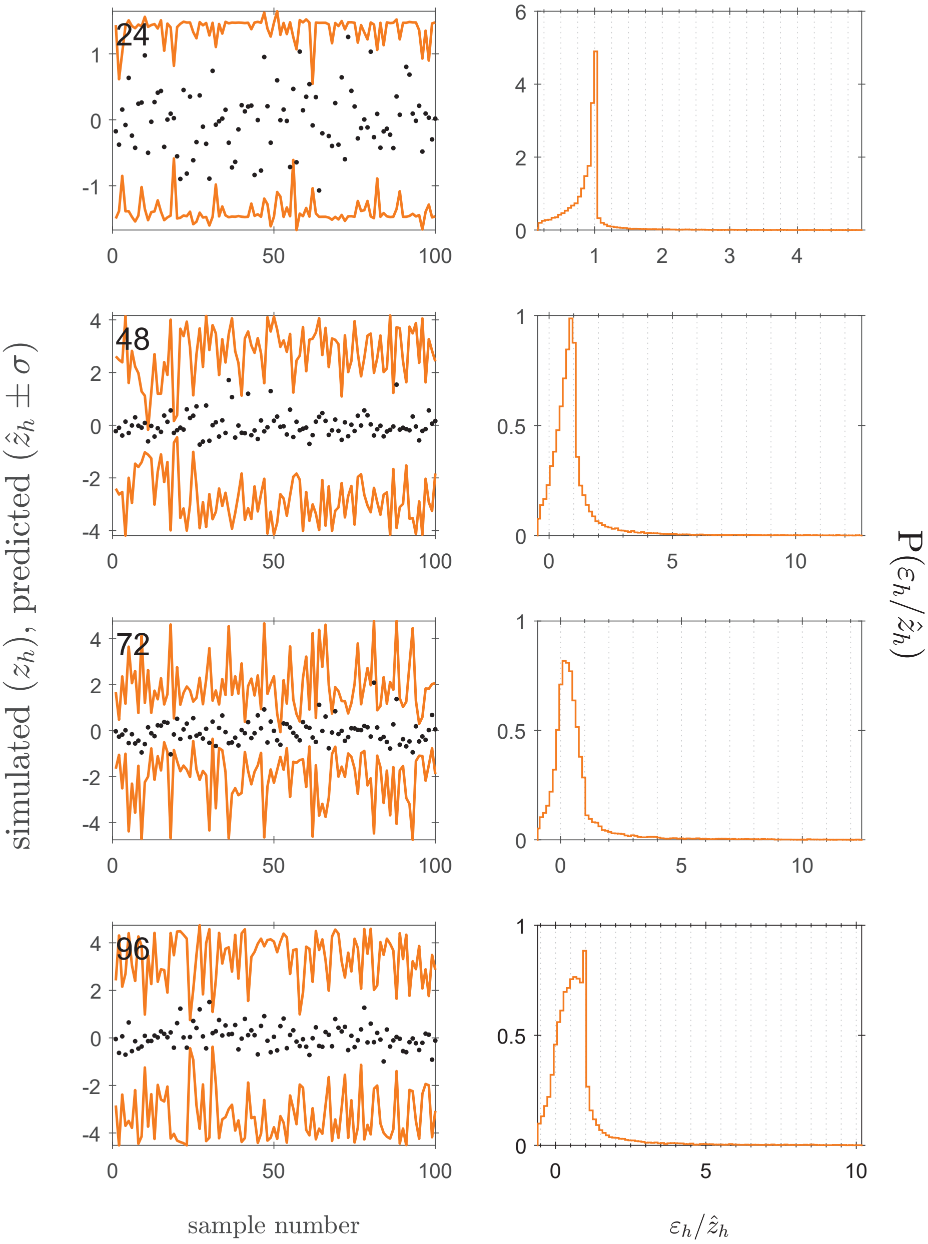


Figure B.13 – *Case 1, Best Fit, ARD kernel – heating. The prediction intervals do not seem to become narrower with more data. These are z-scores, not the original data.*

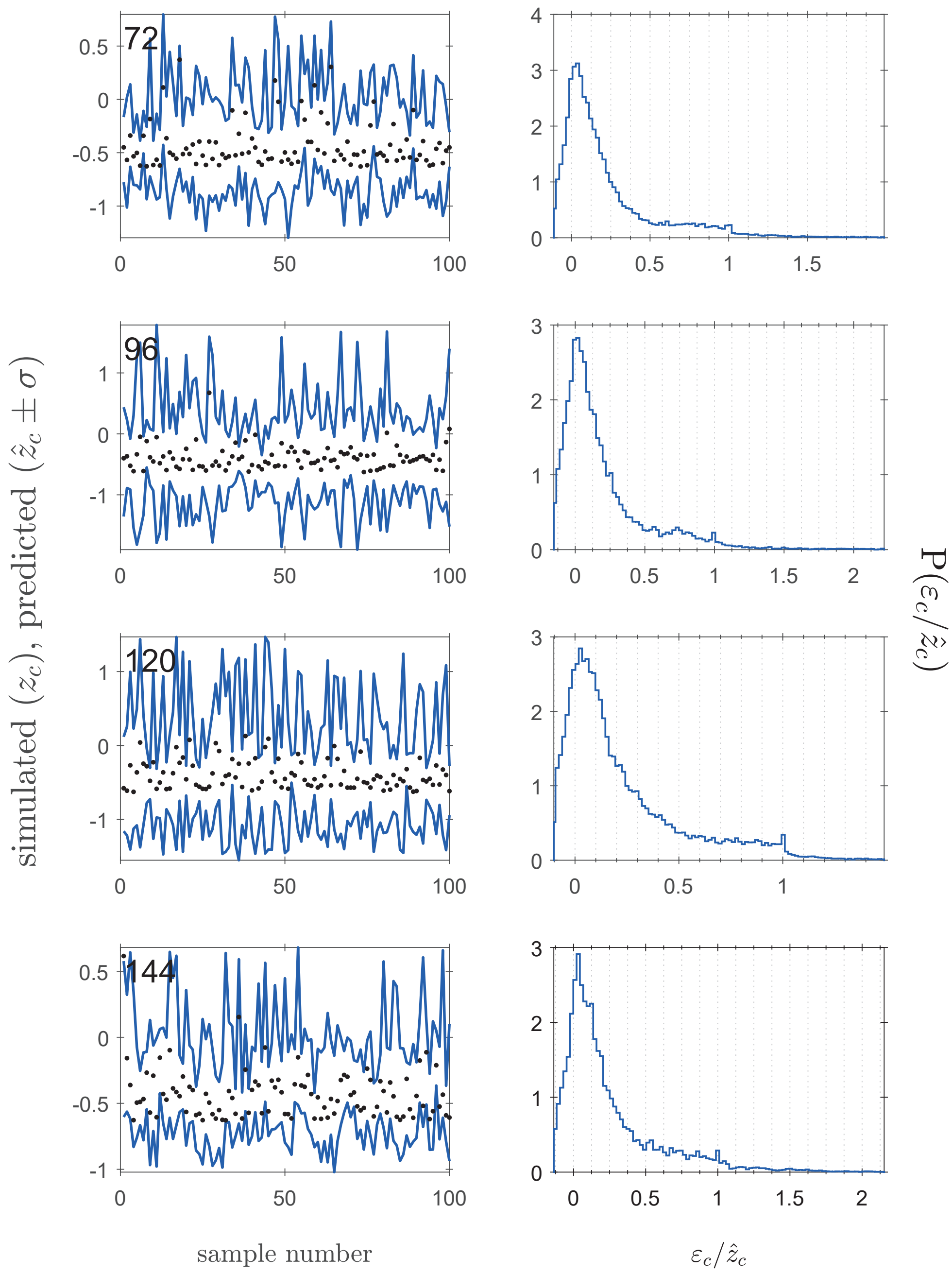


Figure B.14 – *Case 1, Best Fit, ARD kernel – cooling. The prediction intervals do not seem to become narrower with more data. These are z-scores, not the original data.*

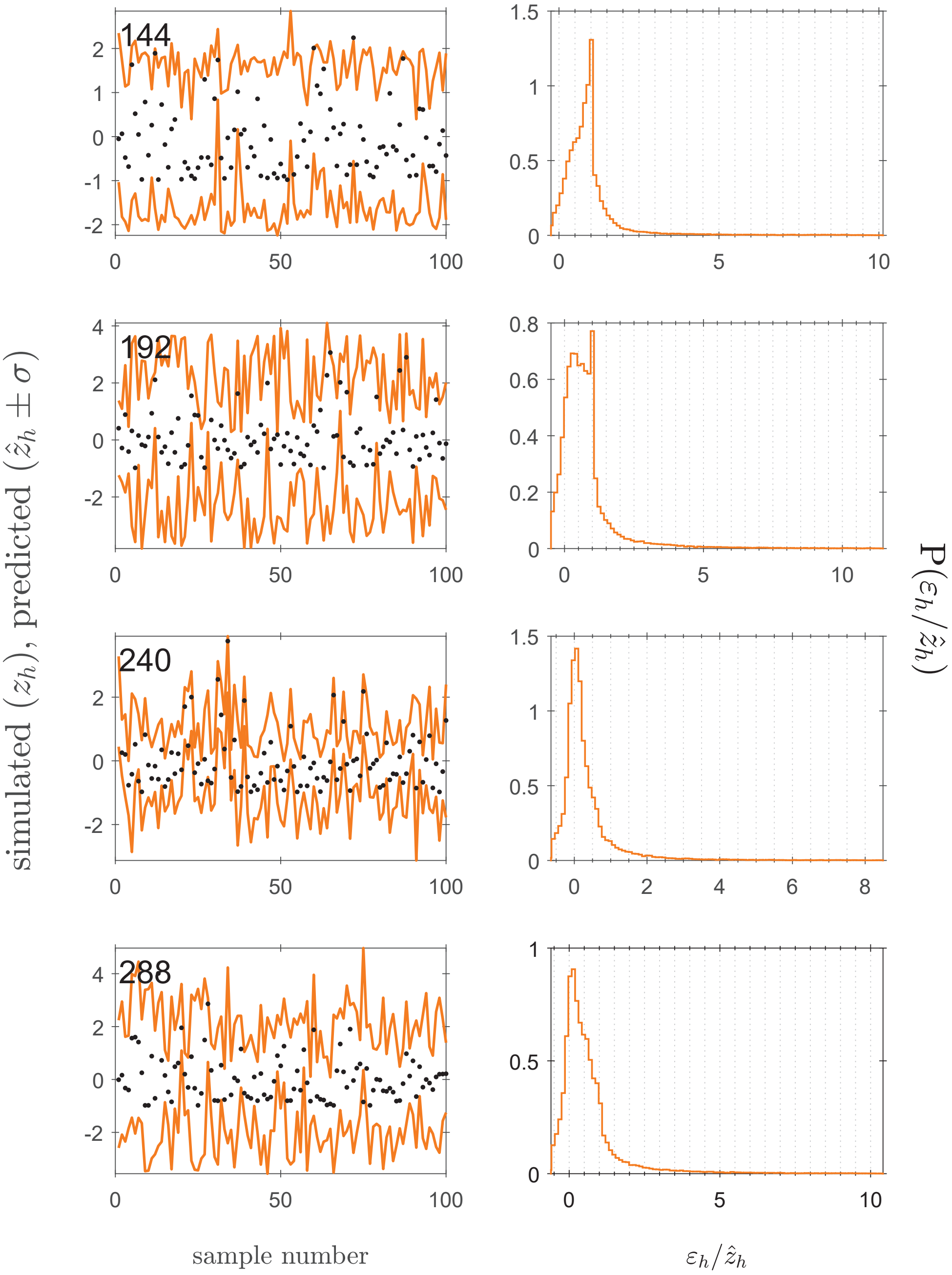


Figure B.15 – *Case 2, Best Fit, ARD kernel – heating. These are z-scores, not the original data.*

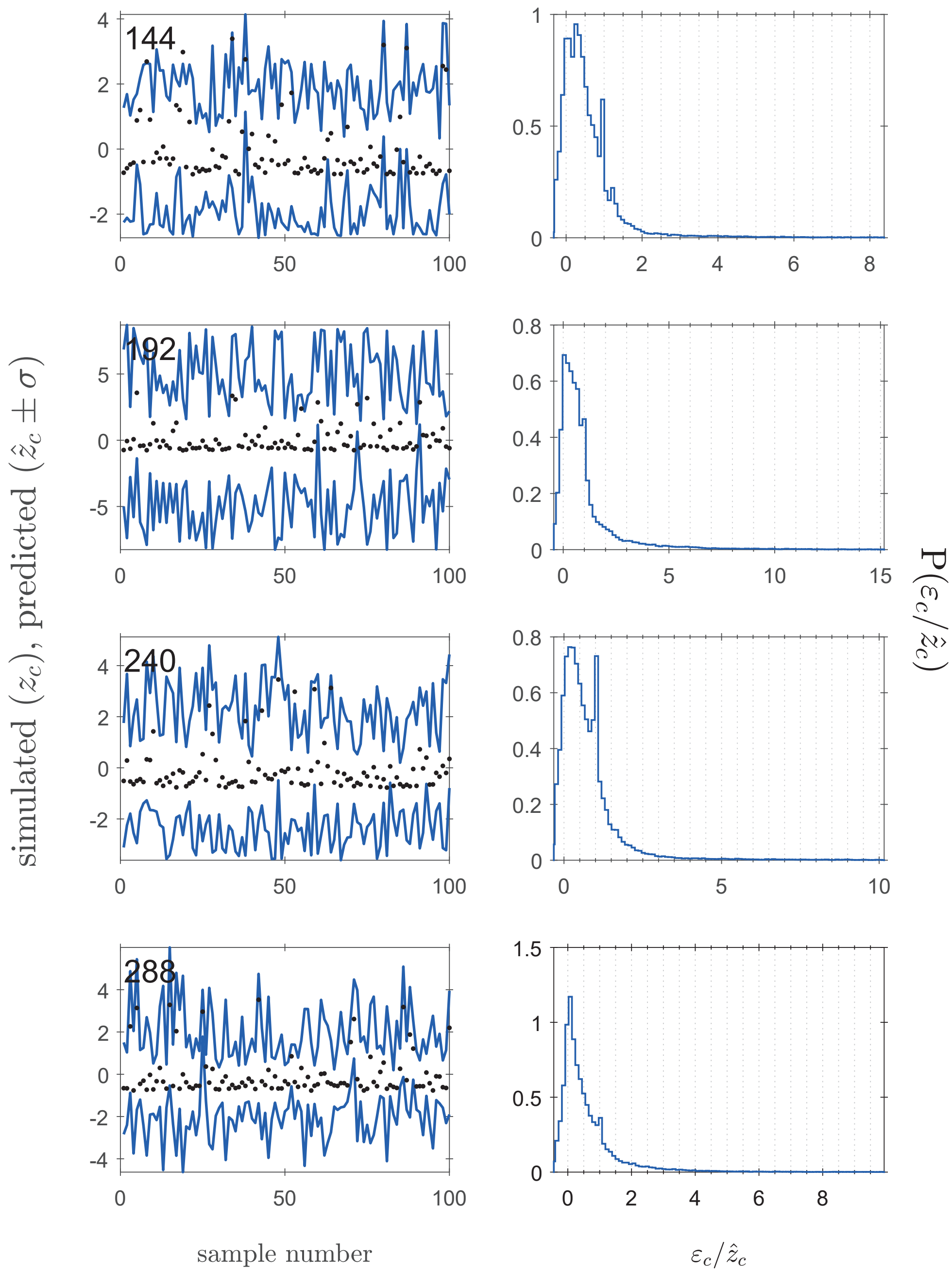


Figure B.16 – *Case 2, Best Fit, ARD kernel – cooling. These are z-scores, not the original data.*

B.3 Regression Inputs: Additional Discussion and Concepts

B.3.1 List of Initial Inputs

In Tables B.1 and B.2 the dash implies that the quantity is dimensionless. Note that the Inter-Quartile Ranges (IQRs) for the solar radiation terms are calculated using the daily sum of radiation values, rather than the raw values. Their means are also calculated using an (empirical) exponential distribution fit to the data, since the distributions of these two variables are strongly non-Normal. The codes presented in these tables are used throughout the thesis to indicate the respective inputs.

Table B.1 – *Initial inputs, codes, and units.*

Group	Quantity	Statistic	Name	Code	Units
Building	U-value	Average	Average U-value of envelope	*uval*	W/m^{2}K
	Thermal Mass	Sum	Sum of thermal storage capacity	*tmass*	MWh/K
	Envelope Ratios	Ratio	Ratio of window area to wall area	*WWR*	—
			Ratio of window area to floor area	*WFR*	
	Massing	Ratio	Form Factor (Volume / Wall Area)	*ff*	—
			Roof Ratio (Roof / Wall Area)	*rr*	
Mixed	Shading	Average	Average sunlit percentage of envelope	*avgsunperc*	%
	Infiltration	Sum	Annual sum of energy gained due to infiltration	*suminfgain*	GWh
			Annual sum of energy lost due to infiltration	*suminfloss*	
Other		Sum	Annual sum of Internal Heat Gain	*sumIHG*	GWh

Table B.2 – *Initial inputs, codes, and units (contd. from Table B.1).*

Group	Quantity	Statistic	Name	Code	Units
CLIMATE	Degree Days	Sum	Annual sum of cooling degree days	*cdd*	°C-day
			Annual sum of heating degree days	*hdd*	
	Dry Bulb Temperature (Hourly)	Average	Annual average of dry bulb temperature	*avgtdb*	°C
		Median	Median dry bulb temperature	*medtdb*	
		IQR	Inter-quartile range of dry bulb temperature	*iqrtdb*	
	Dew Point Temperature (Hourly)	Average	Annual average of dew point temperature	*avgtdb*	°C
		Median	Median dew point temperature	*medtdb*	
		IQR	Inter-quartile range of dew point temperature	*iqrtdb*	
	Global Horizontal Irradiation (Hourly)	Average	Annual average of global horizontal irradiation	*avgghi*	MWh/m^2
		Sum	Annual sum of global horizontal irradiation	*sumghi*	
		IQR	Inter-quartile range of global horizontal irradiation	*iqrghi*	
	Direct Normal Irradiation (Hourly)	Average	Annual average of direct normal irradiation	*avgdni*	MWh/m^2
		Sum	Annual sum of direct normal irradiation	*sumdni*	
		IQR	Inter-quartile range of direct normal irradiation	*iqrdni*	
	Humidity (Hourly)	Average	Annual average of relative humidity	*avgrh*	%
		Median	Median relative humidity	*medrh*	

B.3.2 Dependence and Orthogonality

B.3.2.1 Checking Correlation

While most of the correlations seem to correspond to the physical properties of the input variables, there are a few surprises. For example, it is unsurprising that the means and medians of most variables are highly correlated. We expect that most of the physical quantities under consideration come from populations that, while not necessarily Normal, are at least symmetric about their mean. The solar variables do not conform to this rule, but in their case we find a strong correlation between the sum and means of the quantities. The average sunlit percentage of the facade, which is a reflection both of sunlit hours and self-shading, shows virtually no correlation with any other solar variable. This makes it a strong candidate for inclusion. There is correlation across categories as well, for example, that between average TDB and sum of Global Horizontal Irradiation (GHI). This is understandable from a physical perspective (sunnier climates are likely to be warmer) , and has been exploited in the time series methods (Chapter 3) for reconstructing daily solar profiles using only the daily means of GHI and TDB.

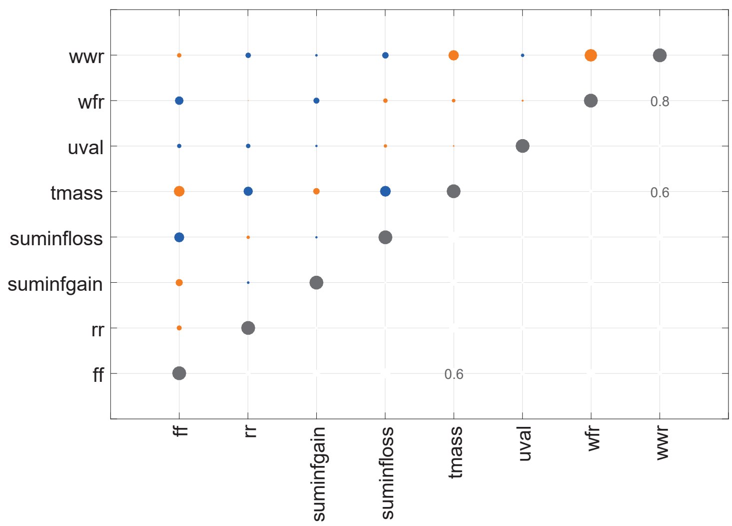


Figure B.17 – *Correlation coefficients for Case 2. The behaviour is similar to that from Case 1 (Figure 4.3). Lighter grey dots indicate positive correlation, while the darker grey denote negative. The black dots on the diagonal indicate an auto-correlation of 1.*

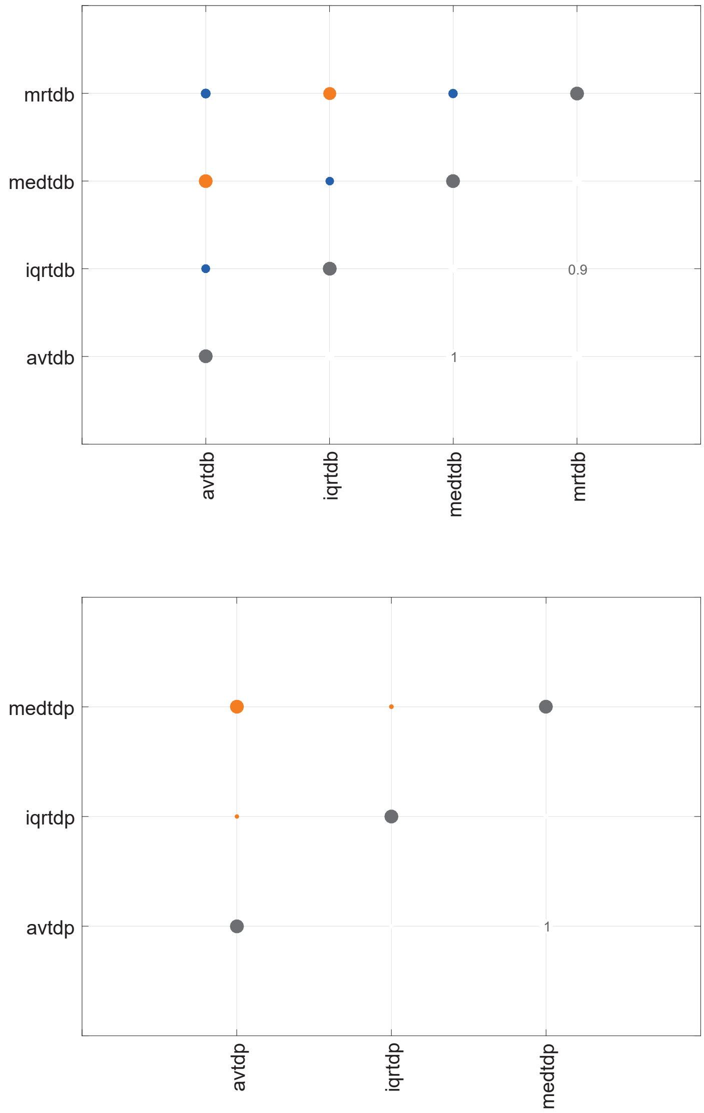
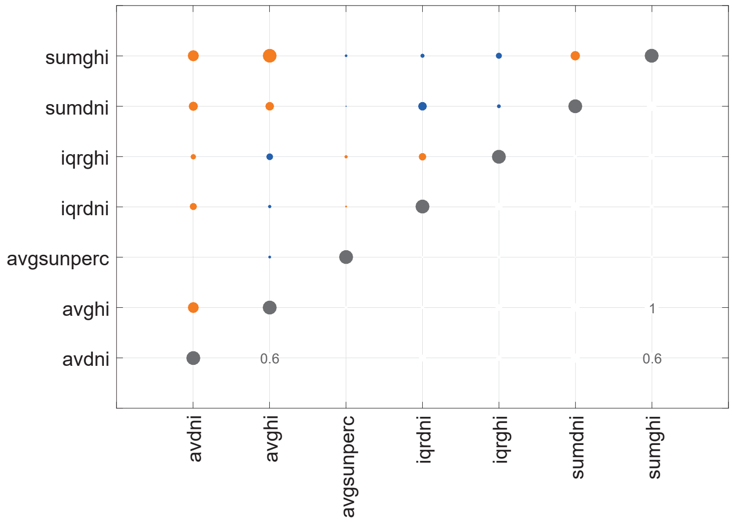


Figure B.18 – *Correlation coefficients for temperature-based inputs, TDB and Dew Point Temperature (TDP). See Table 4.1 for the list of codes. Orange dots indicate positive correlation, while the blue denote negative. The size of each dot indicates the strength of correlation (larger dot equals stronger correlation). The black dots on the diagonal indicate correlation of 1.*



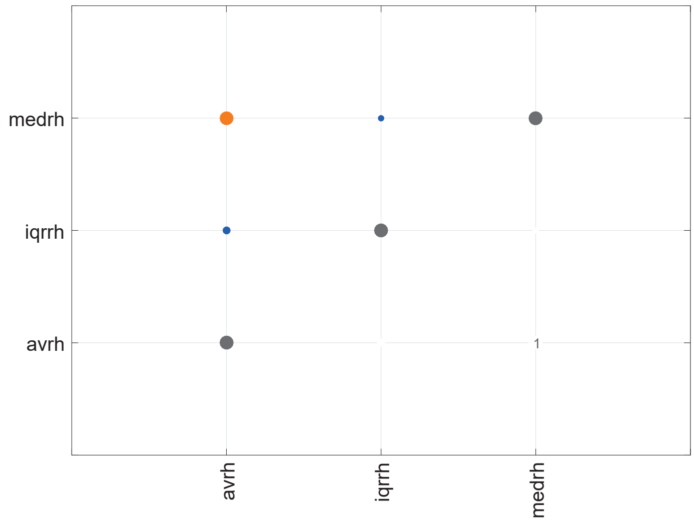


Figure B.19 – *Correlation coefficients for RH- and solar-based inputs.*

B.3.2.2 Principal Component Analysis

Originally proposed by Pearson (1901) but developed in its modern form by Hotelling (1933), Principal Component Analysis (PCA) is a technique where dimensionality reduction is achieved by estimating a new orthogonal basis space – the space defined by the Principal Components (PCs) – where the new basis vectors explain as much of the variance in the original data as possible. Christensen (1991) prefers to explain PCA using predictive terminology, i.e., that the Principal Components (PCs) are the Best Linear Unbiased Predictors (BLUPs) of the original variables. He further describes PCA in terms of the original problem framed by Hotelling (1933). For PCA to be properly useful, the number of PCs should be less than the number of original predictor variables. However, the orthogonality of the PCs is also useful. A completely orthogonal set of basis vectors is the most efficient way of representing a space.

The PCs are defined sequentially, in decreasing order of variance, and progressively individually explain *less and less* of the original variance. See Christensen (1991) for a derivation of why the first r linear PCs "have the maximum capability to predict [an original set x] among all sets of r linear combinations of [x] ." Say we are using two Principal Components (PCs) ($\mathbb{R}^2$) to define a subset of the original predictors $\mathbb{R}^q$. Then the first PC is

$$t_{1,j} = a_1 \cdot x_{1,j} + a_2 \cdot x_{2,j} + \cdots + a_{q,j} \cdot x_{q,j} = \sum_{i=1}^{q} a_i \cdot x_i, \tag{B.12}$$

where t_j is the value of the new PCs at some point $j = 1, \ldots, n$, and $a_1, a_2, \ldots, a_q$ are coefficients of the linear combination of the original variables (also called weights or loadings).

The second PC is

$$t_{2,j} = b_1 \cdot x_{1,j} + b_2 \cdot x_{2,j} + \cdots + b_{q,j} \cdot x_{q,j} = \sum_{i=1}^{q} b_i \cdot x_i, \tag{B.13}$$

where $b_1, b_2, \ldots, b_q$ is a different set of coefficients from $a_1, a_2, \ldots, a_q$. Subsequent PCs follow the same general idea. The coefficients of the second PC maximise the variance explained in the data matrix after the first PC has been subtracted from it.

Christensen (ibidem) warns that the Principal Components (PCs) are subject to the vagaries of scale, for two reasons. First, the variance of any individual component can be changed by multiplying the variable with a constant, e.g., $\mathrm{Var}(aY) = a^2\mathrm{Var}(Y)$, where Y is the variable and a is a constant. Quantities that happen to not have large absolute variance in the sample will be ignored. The author suggests standardising the variables, which we do in our analyses by taking z-scores.

Christensen (ibidem) ends his explanation of PCA with two warnings. For one, since Principal Components (PCs) are based on linear combinations and predictors, they might miss non-linear structures in the model entirely. In our case we suspect that the responses are, at best, piecewise linear. Secondly, PCs are calculated from a specific sample. Depending on the sampling scheme, the variances of different variables may change from sample to sample. The optimal data reduction, then, relies on a representative sample. We contend that, mathematical considerations aside, dimensionality should only be reduced if it makes *physical sense*, so we do not advise using PCs when their physical meaning is not clear. For example, what would a PC made up of Window-to-Wall Ratio and Thermal Mass mean?

B.3.3 Scaling Options for Inputs and Outputs

McCullagh and Nelder (1983) say that a good scale or scaling parameter has to "combine constant variance and approximate Normality of errors with additivity of systematic effects". In theory any transformation of the form $\hat{x} = f(x)$ could be carried out on the original data, where $f(\cdot)$ is some function. Common transforms include a Probability Density Function (PDF) mapping the original data to some distribution; domain transforms like Fourier and Laplace; and the power transforms family. A widely used sub-family of power transforms is the one-parameter Box-Cox family

$$\hat{y}_i = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases} \tag{B.14}$$

where λ is a scaling parameter. Transforms invariably complicate physical interpretation of the model, even if a good enough transformation parameter λ is found, i.e., one that gives a Normal distribution.

B.4 Simulation Details

For the original Energy Plus input files (IDF) used in the thesis, the user is directed to the EPFL archive : infoscience.epfl.ch[1]. The IDF files are included as 'supplementary documents' in the archive, like the MATLAB and R code used throughout this thesis. The link to a GIT repository for the code will also be found there. The Energy Plus files for the United States Department of Energy (USDOE) buildings are available freely online.

Table B.3 – *The commercial reference buildings from Deru, Field et al. (2011). The codes reflect the era (first digit) and the building (next two digits). The 'Codes' and 'U-value' columns are split according to era: 1 – pre-1980s, 2 – Post-1980s, 3 – New Construction.*

Codes			Building Type	Name	U-value $[W/m^2K]$		
1	2	3			1	2	3
D101	D201	D301	Office	Large	2.28	2.46	3.5
D102	D202	D302		Medium	2.46	1.76	1.44
D103	D203	D303		Small	1.57	3.45	3.44
D104	D204	D304	Warehouse		1.13	1.95	6.06
D105	D205	D305	Retail	Stand-Alone	1.48	1.69	1.65
D106	D206	D306		Strip-Mall	2.23	2.56	2.01
D109	D209	D309		Supermarket	0.934	1.11	1.07
D107	D207	D307	School	Primary	1.61	1.81	1.43
D108	D208	D308		Secondary	1.66	1.96	1.48
D110	D210	D310	Restaurant	Quick-Service	1.19	4.35	3.32
D111	D211	D311		Full-Service	1.12	4.52	1.12
D112	D212	D312	Health	Hospital	1.55	2.03	1.55
D113	D213	D313		Outpatient	1.53	2.17	1.23
D114	D214	D314	Hotel	Small	2.76	3.51	1.24
D115	D215	D315		Large	1.99	2.51	1.87
D116	D216	D316	Home	Apartment	1.66	2.46	3.56

[1] Search by author name or thesis title.

Table B.4 – *The renovation cases of a single-family home analysed in this thesis. The 'old' codes correspond to the nomenclature in Chinazzo (2014), while the 'new' codes are used in this thesis. We added two extra variations of the base case, changing the external walls to get worse U-values than the base case.*

Old Code	New Code	U-value	Description
BC00	G000C00	1.70	Base Case
–	G000C01[a]	2.65	
–	G000C02[b]	2.85	
RC01	G000C03	1.09	External wall insulation
RC02	G000C04	1.04	
RC03	G000C05	0.98	
RC04	G000C06	1.08	Internal wall insulation
RC05	G000C07	1.04	
RC06	G000C08	0.98	
RC07	G000C09	1.44	External roof insulation
RC08	G000C10	1.41	
RC09	G000C11	1.39	
RC10	G000C12	1.44	Internal roof insulation
RC11	G000C13	1.41	
RC12	G000C14	1.39	
RC13	G000C15	1.70	Floor insulation
RC14	G000C16	1.70	
RC15	G000C17	1.70	
RC16	G000C18	1.41	Window substitution
RC17	G000C19	1.39	
RC18	G000C20	1.32	
RC19	G000C21	1.70	Shading Systems
RC20	G000C22	1.70	
~~RC21~~	~~G000C23~~	–	~~Phase-Change Material~~[c]
RC22	G000C24	1.70	Envelope airtightness

[a] Masonry-only wall
[b] Brick-only wall
[c] Phase-Change Materials are not considered in our analyses.

Figure B.20 – *Screenshot of the home in Geneva (case 1), taken from DesignBuilder®.*

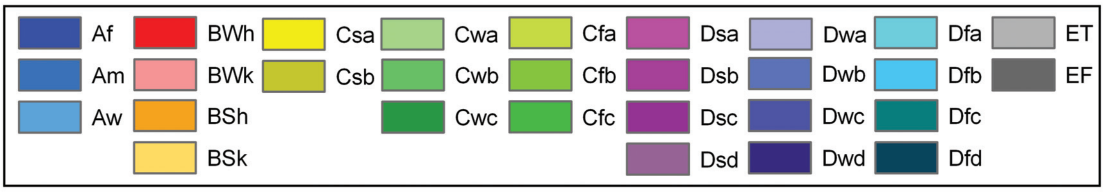


Figure B.21 – *Legend for the Koeppen-Geiger colour scheme, from Peel, Finlayson et al. (2007). This scheme is used in Figures B.22 to B.24.*

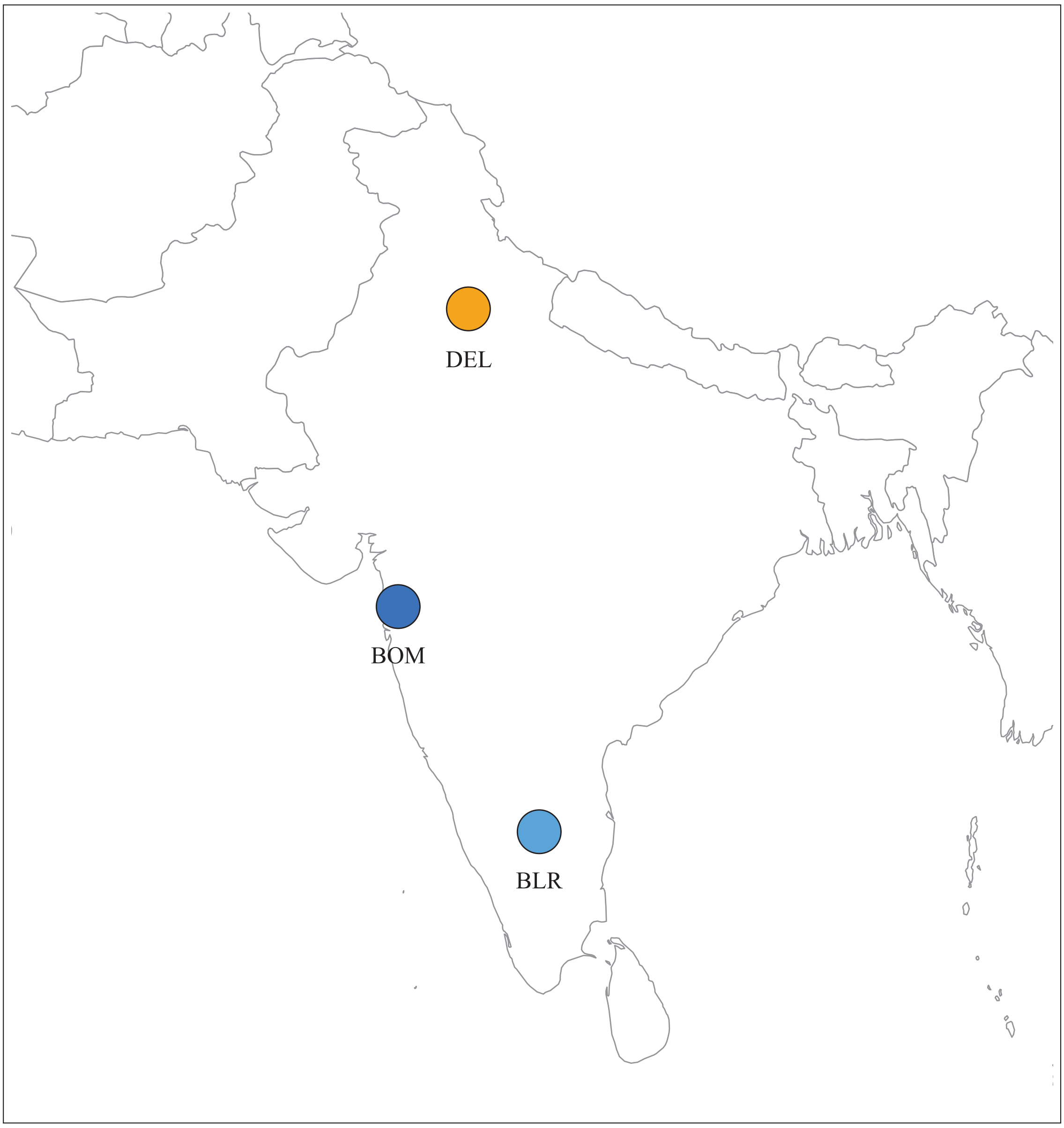


Figure B.22 – *Stations in the Indian subcontinent. Each colour represents a Koeppen-Geiger climate zone, with the scheme corresponding to that of Peel, Finlayson et al. (2007), reproduced in Figures 2.1 and B.21.*

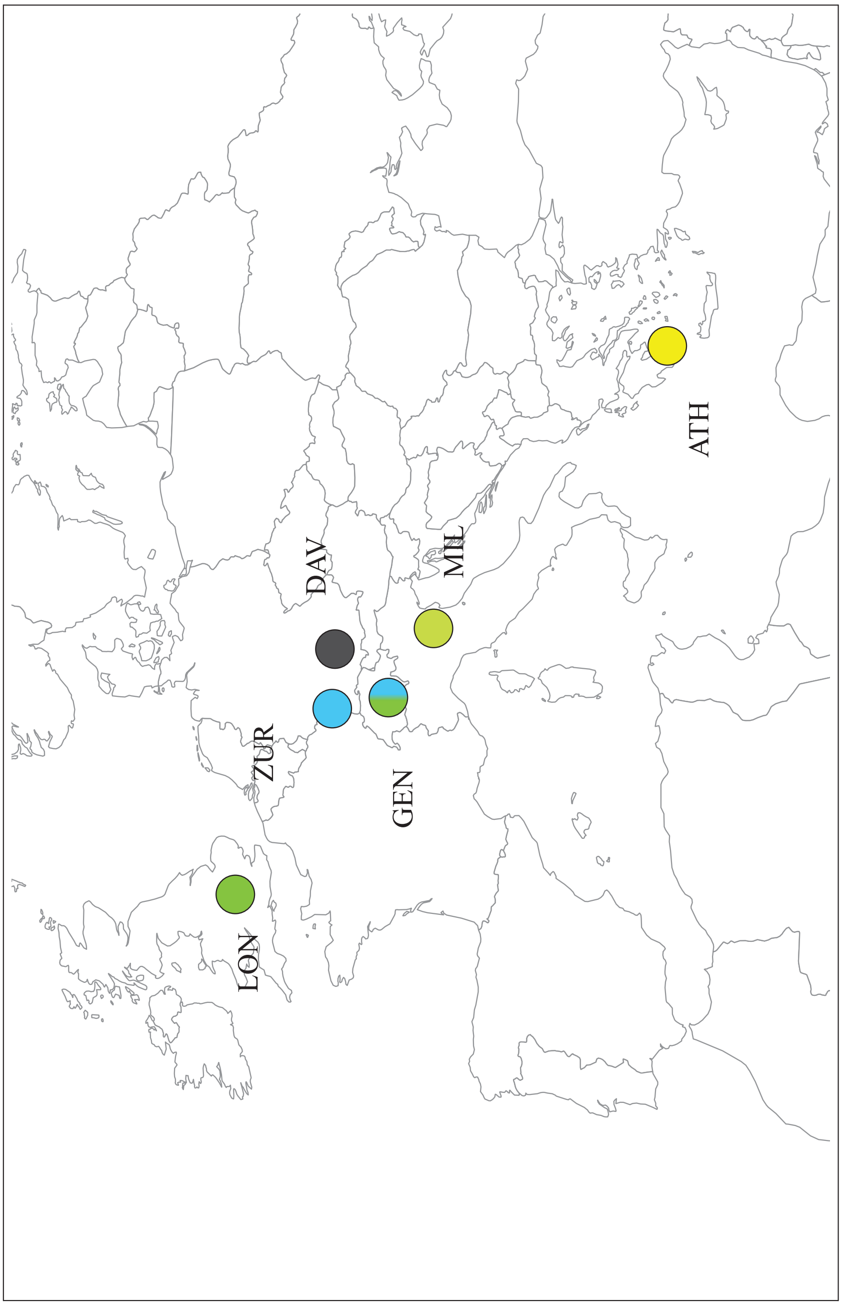


Figure B.23 – *Stations in Europe. Circles with more than one colour are for those climates where the divisions are too close to select one. Colours based on Peel, Finlayson et al. (2007), reproduced in Figures 2.1 and B.21.*

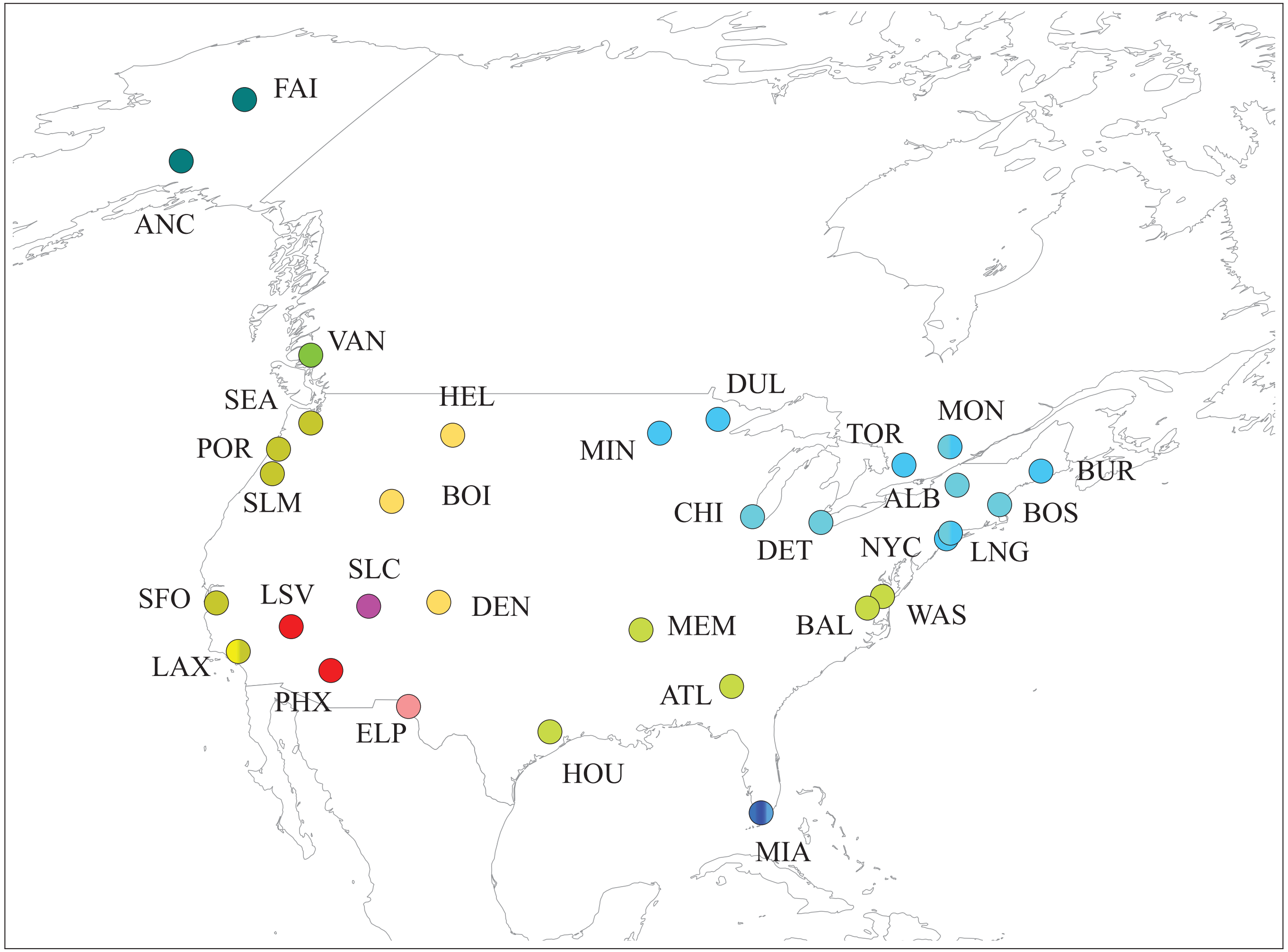


Figure B.24 – *Stations in North America. The dots are placed as close to the actual location of the station as possible. Colours based on Peel, Finlayson et al. (2007), reproduced in Figures 2.1 and B.21.*

Table B.5 – *Details of the climates/stations used in the regression examples from Chapter 4.*

City Name	Station	Code	WMO No.	Longitude (°)	Latitude (°)	ASHRAE Zone	Koeppen-Geiger
ALBANY	Albany Intl. Arpt.	ALB	723650	-106.62	35.04	5A/6A	Dfa
ANCHORAGE	Ted Stevens Intl. Arpt.	ANC	702730	-150	61.18	7	Dfc
ATHENS	Athene	ATH	167160	23.73	37.9	3B/4B	Csa
ATLANTA	Dekalb Peachtree Fulton Co Arpt Brow Hartsfield Jackson Intl. Arpt.	ATL_DEK ATL_FUL ATL_JAC	722196 722195 722190	-84.43 -84.52 -84.3	33.63 33.77 33.87	3A/4A	Cfa
BALTIMORE	Andrews Air Force Base Baltimore Washington Arpt.	BAL_AFB BAL_WAS	745940 724060	-76.87 -76.68	38.82 39.17	3A/4A	Cfa
BANGALORE	Bengaluru	BLR	432950	77.58	12.97	2B	Aw
BOISE	Air Terminal	BOI	726810	72.82	18.9	5B/6B	BSk
BOMBAY	Mumbai	BOM	430030	-71.02	42.37	1B	Am
BOSTON	Logan Arpt.	BOS	725090	88.45	22.65	5A	Dfa
BURLINGTON	Intl. Arpt. Municipal Arpt. Piers	BUR_AIR BUR_MUN BUR_PIE	726170 725455 714370	-87.75 -87.92 9.844	41.78 41.98 46.813	5A/6A	Dfb
CHICAGO	Midway Arpt. O'Hare Arpt.	CHI_MDW CHI_OHR	725340 725300	77.2 -104.65	28.58 39.83	5A	Dfa

Table B.6 – *Details of the climates/stations used in the regression examples from Chapter 4 (continued from Table B.5).*

City Name	Station	Code	WMO No.	Longitude (°)	Latitude (°)	ASHRAE Zone	Koeppen-Geiger
DAVOS	Davos	DAV	067840	-104.65	39.83	7	ET
DELHI	New Delhi	DEL	421820	-83.35	42.22	1B	BSh
DENVER	Stapleton Arpt. Golden	DEN_AIR DEN_GOL	725650 724666	-83.53 6.13	42.23 46.25	5B/6B/7B	BSk
DETROIT	City Metropolitan Arpt. Willow Run Arpt.	DET_CTY DET_MET DET_WIL	725375 725370 725376	-95.37 -95.55 -95.28	30 30.07 29.65	5A	Dfa
DULUTH	Harbor	DUL_HRB	727456	-118.4	33.93	7A	Dfb
EL PASO	Intl. Arpt.	ELP	722700	-73.1	40.78	3B	BWk
FAIRBANKS	Intl. Arpt. Eielson	FAI_AIR FAI_EIE	702610 702650	-0.18 -115.15	51.15 36.08	8	Dfc
GENEVA	Cointrin Arpt.	GEN	067000	-80.3	25.82	4/5 A/B	Cfb/Dfb
HELENA	Regional Arpt.	HEL	727720	-80.43	25.65	6B	BSk
HOUSTON	Bush Arpt. DW Hook Arpt. William P Hobby Arpt.	HOU_BUS HOU_HOO HOU_WIL	722430 722429 722435	-95.37 -95.55 -95.28	30.00 30.07 29.65	2A	Cfa
LOS ANGELES	LosAngeles	LAX	722950	-93.35	45.07	3B	Csa/Csb
LONG ISLAND	LongIsland	LNG	725035	-73.75	45.47	5A/6A	Dfa
LONDON	Gatwick	LON_LGW	037760	-74.03	45.68	4A	Cfb
LAS VEGAS	McCarran Arpt.	LSV	723860	-73.62	45.5	3B/4B/5B	BWh

Table B.7 – *Details of the climates/stations used in the regression examples from Chapter 4 (continued from Tables B.5 and B.6).*

City Name	Station	Code	WMO No.	Longitude (°)	Latitude (°)	ASHRAE Zone	Koeppen-Geiger
Memphis	Intl. Arpt.	MEM	723340	-73.97	40.78	3A/4A	Cfa
Miami	Intl. Arpt.	MIA_AIR	722020	-73.8	40.65	1A	Am/Af/Aw
	Kendall Executive	MIA_EXE	722029	-73.88	40.78		
	Opa Locka	MIA_OPA	722024	-70.3	43.65		
Milan	Linate	MIL_LIN	160800	46.8	24.7	4A	Cfa
	Malpensa	MIL_MAL	160660	-111.97	40.77		
Minneapolis	St Paul Intl. Arpt.	MIN_AIR	726580	-122.25	47.68	6A/7A	Dfb
	Crystal	MIN_CRY	726575	-122.32	47.47		
Montreal	Jean Brebeuf	MON_JBF	071612	-79.63	43.67	6A	Dfa/Dfb
	Mirabel Arpt.	MON_YMX	716278	-123.17	49.18		
	Pierre E. Trudeau Intl. Arpt.	MON_YUL	716270	-77.03	38.87		
New York	Central Park	NYC_CPR	725033	-77.47	38.98	4A/5A	Dfa/Dfb
	John F Kennedy Arpt.	NYC_JFK	744860	8.536	47.48		
	La Guardia Arpt.	NYC_LAG	725030	8.567	47.378		
Phoenix	Deer Valley	PHX_DEE	722784	-73.15	44.47	2B/3B/4B	BWh
	Sky Harbor Intl Arpt.	PHX_SKY	722780	-79.8	43.3		
Portland	Jetport	POR	726060	-91.11	40.78	4C	Csb
Salt Lake City	SaltLakeCity	SAL	725720	-121.93	37.36	5B/6B	Dsa

Table B.8 – *Details of the climates/stations used in the regression examples from Chapter 4 (continued from Tables B.5 to B.7).*

City Name	Station	Code	WMO No.	Longitude (°)	Latitude (°)	ASHRAE Zone	Koeppen-Geiger
Seattle	Boeing Field Arpt.	SEA_BOE	727935	-121.81	37.33	4C	Csb
	Tacoma Arpt.	SEA_TAC	727930	-122.4	37.62		
San Francisco	Intl. Arpt.	SFO_AIR	724940	-122.12	37.67	3C	Csb
	Hayward Air Terminal	SFO_HAY	724935	-122.22	37.72		
	Oakland Metro. Arpt.	SFO_OAK	724930	-147.85	64.82		
	SanJose Intl.	SFO_SJO	724945	-147.1	64.65		
	SanJose/Reid/Hillv	SFO_SJR	724946	-92.22	46.83		
Salem	Mcnary Field	SLM	726940	-92.08	46.76	4C	Csb
Toronto	Pearson Intl. Arpt.	TOR	716240	-111.97	46.6	6A	Dfb
Vancouver		VAN	718920	-89.98	35.07	4C/6A	Cfb
Washington DC	Dulles Arpt.	WAS_DUL	724030	-112.08	33.68	3A/4A	Cfa
	Reagan Arpt.	WAS_RGN	724050	-111.98	33.45		
Zurich	Kloten	ZUR_KLO	066700	-106.5	31.77	4/5 A/B	Dfb
	Affoltern	ZUR_REH	066600	-116.21	43.62		

Bibliography

3TIER (2011). *The Risks of Using TMY Data*.

al-Abbadi, N (2002). *NASA Remote Sensing Validation Data: Saudi Arabia*.

Abdi, H and LJ Williams (2010). 'Principal Component Analysis'. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4, pages 433–459.

Acharya, PK (1928). *Indian Architecture According to Mānasāra-śilpaśāstra*. en. Patna : Indian India.

Ackermann, M (2010). *Cool Comfort*. en. Smithsonian.

Adelard, L, H Boyer, F Garde and JC Gatina (2000). 'A detailed weather data generator for building simulations'. In: *Energy and Buildings* 31.1, pages 75–88.

Adelard, L, TA Mara, H Boyer and JC Gatina (1999). 'Elaboration of a new tool for weather data sequences generation'. In: *Proceedings of BS 1999*. Kyoto: IBPSA.

Agarwal, M, P Rastogi, M Peltier, L Pastore and M Andersen (2016). 'Examining Building Design Decisions Under Long Term Weather Variability and Microclimate Effects: A Case Based Exploratory Study'. In: Los Angeles, USA.

Aguiar, R, S Camelo and H Gonçalves (1999). 'Assessing the value of typical meteorological years built from observed and from synthetic data for building thermal simulation'. In: *... ob Building Simulation'99 in Kyoto*.

Aguiar, R and M Collares-Pereira (1992). 'TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation'. In: *Solar Energy* 49.3, pages 167–174.

Aguiar, R, M Collares-Pereira and J Conde (1988). 'Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices'. In: *Solar Energy* 40.3, pages 269–279.

Aguiar, R, M Oliveira and H Gonçalves (2002). 'Climate change impacts on the thermal performance of Portuguese buildings. Results of the SIAM study'. In: *Building Services Engineering Research and Technology* 23.4, pages 223–231.

Aïssani, A, A Chateauneuf and JP Fontaine (2015). 'Reliability based design optimization of insulation systems considering climate change and workmanship uncertainties'. In:

Albers, RAW, PR Bosch, B Blocken, AAJF van den Dobbelsteen, LWA van Hove, TJM Spit, F van de Ven, T van Hooff and V Rovers (2015). 'Overview of challenges and achievements in the climate adaptation of cities and in the Climate Proof Cities program'. In: *Building and Environment*. Special Issue: Climate adaptation in cities 83, pages 1–10.

Alkhalaf, A and H Kraus (1993). *Energy balance equivalents to the Koeppen-Geiger climatic regions*. Dümmler.

Alsaadani, S and CB de Souza (2012). 'The Social Component Of Building Performance Simulation: Understanding Architects'. In: edited by J Wright and M Cook. IBPSA-England, pages 332–339.

Amiri, SS, M Mottahedi and S Asadi (2015). 'Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S.' In: *Energy and Buildings* 109, pages 209–216.

AMS (2015). *Meteorology Glossary*.

Andersen, M, S Kleindienst, L Yi, J Lee, M Bodart and B Cutler (2008). 'An intuitive daylighting performance analysis and optimization approach'. In: *Building Research & Information* 36.6, pages 593–607.

Ansuini, R, A Giretti and M Lemma (2012). 'A probabilistic approach to decision making in conceptual design'. In: edited by J Wright and M Cook. IBPSA-England, pages 213–220.

Argiriou, A, S Lykoudis, S Kontoyiannidis, CA Balaras, D Asimakopoulos, M Petrakis and P Kassomenos (1999). 'Comparison of methodologies for tmy generation using 20 years data for Athens, Greece'. In: *Solar Energy* 66.1, pages 33–45.

Asadi, S, SS Amiri and M Mottahedi (2014). 'On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design'. In: *Energy and Buildings* 85, pages 246–255.

Ascione, F, N Bianco, R De Masi, G Mauro and G Vanoli (2015). 'Design of the Building Envelope: A Novel Multi-Objective Approach for the Optimization of Energy Performance and Thermal Comfort'. en. In: *Sustainability* 7.8, pages 10809–10836.

ASHRAE (2013). *Standard 169-2013 – Climate Data for Building Design Standards*. Standard D-86519. Atlanta, GA.

ASHRAE, AIA, IESNA, USGBC and USDOE (2004). *ASHRAE Advanced Energy Design Guide for Small Office Buildings*. Technical report 1931862559. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

– (2011). *ASHRAE Advanced Energy Design Guide for Small to Medium Office Buildings*. Technical report 9781936504053. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

ASHRAE and ANSI (2010). *Standard 55-2010, Thermal Environmental Conditions for Human Occupancy*. Standard 55-2010. Atlanta, GA.

– (2011). *Standard 140-2011 – Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs (ANSI Approved)*. Standard 140-2011. Atlanta, GA.

ASHRAE, ANSI and IESNA (2010). *Standard 90.1-2010: Energy Efficient Design of New Buildings Except Low-Rise Residential Buildings*. Standard 90.1-2010. Atlanta, GA.

ASHRAE, IESNA, USGBC and ANSI (2011). *Standard 189.1-2011: Design of High-Performance, Green Buildings Except Low-Rise Residential Buildings*. Standard 189.1-2011. Atlanta, GA.

Athienitis, AK (1989). 'A computer method for systematic sensitivity analysis of building thermal networks'. In: *Building and Environment* 24.2, pages 163–168.

Athienitis, AK, M Chandrashekar and HF Sullivan (1985). 'Modelling and analysis of thermal networks through subnetworks for multizone passive solar buildings'. In: *Applied Mathematical Modelling* 9.2, pages 109–116.

Athienitis, AK, HF Sullivan and KGT Hollands (1986). 'Analytical model, sensitivity analysis, and algorithm for temperature swings in direct gain rooms'. In: *Solar Energy* 36.4, pages 303–312.

Athienitis, A, H Sullivan and K Hollands (1987). 'Discrete fourier series models for building auxiliary energy loads based on network formulation techniques'. en. In: *Solar Energy* 39.3, pages 203–210.

Athienitis, AK and M Santamouris (2002). *Thermal analysis and design of passive solar buildings*. London: James & James.

Attia, S, E Gratia, A De Herde and JL Hensen (2012). 'Simulation-based decision support tool for early stages of zero-energy building design'. In: *Energy and Buildings* 49, pages 2–15.

Attia, S, M Hamdy, W O'Brien and S Carlucci (2013). 'Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design'. In: *Energy and Buildings* 60, pages 110–124.

Attia, S, JL Hensen, L Beltrán and A De Herde (2012). 'Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs'. In: *Journal of Building Performance Simulation* 5.3, pages 155–169.

Augenbroe, G (1992). 'Integrated building performance evaluation in the early design stages'. In: *Building and Environment* 27.2, pages 149–161.

Banham, R (1984). *Architecture of the Well-Tempered Environment*. 2nd. Chicago: The University of Chicago Press.

Bansal, NK, G Hauser and G Minke (1994). *Passive building design: a handbook of natural climatic control*. Elsevier Science B.V.

Bansal, NK and G Minke (1995). *Climatic zones and rural housing in India: German-Indian cooperation in scientific research and technological development*. Scientific series of the International Bureau, v. 35. Forschungszentrum Jülich GmbH, Zentralbibliothek.

Belcher, SE, JN Hacker and DS Powell (2005). 'Constructing design weather data for future climates'. en. In: *Building Services Engineering Research and Technology* 26.1, pages 49–61.

Bell, B (1970). 'The Oldest Records of the Nile Floods'. In: *The Geographical Journal* 136.4, pages 569–573.

Bhandari, M, S Shrestha and J New (2012). 'Evaluation of weather datasets for building energy simulation'. In: *Energy and Buildings* 49, pages 109–118.

Blight, TS and DA Coley (2013). ‘Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house dwellings’. In: *Energy and Buildings* 66, pages 183–192.

Bloomfield, P (2000). *Fourier analysis of time series: an introduction*. 2nd ed. Wiley series in probability and statistics. Applied probability and statistics section. New York: Wiley.

Boland, J (1984). ‘Time Series Modelling of Solar Radiation’. In:

– (1995). ‘Time-series analysis of climatic variables’. In: *Solar Energy* 55.5, pages 377–388.

– (1997). ‘Simplifying the solution of the differential equations which describe heat flows in domestic dwellings’. In: *Building and Environment* 32.5, pages 479–484.

Booth, AT and R Choudhary (2013). ‘Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal’. In: *Energy and Buildings* 64, pages 292–308.

Booth, AT, R Choudhary and DJ Spiegelhalter (2012). ‘Handling uncertainty in housing stock models’. In: *Building and Environment* 48, pages 35–47.

Box, GEP, GM Jenkins and GC Reinsel (2008). *Time Series Analysis: Forecasting and Control, Fourth Edition*. en. John Wiley & Sons.

Breesch, H and A Janssens (2005). ‘Building simulation to predict the performances of natural night ventilation: uncertainty and sensitivity analysis’. In: *Proceedings of BS 2005*. Montréal.

Brohus, H, C Frier, P Heiselberg and F Haghighat (2012). ‘Quantification of uncertainty in predicting building energy consumption: A stochastic approach’. In: *Energy and Buildings* 55, pages 127–140.

Bulut, H (2010). ‘Generation of representative solar radiation data for Aegean Region of Turkey’. In: *International Journal of Physical Sciences* 5.7, pages 1124–1131.

Burhenne, S, O Tsvetkova, D Jacob, GP Henze and A Wagner (2013). ‘Uncertainty quantification for combined building performance and cost-benefit analyses’. In: *Building and Environment* 62, pages 143–154.

Burkhart, MC, Y Heo and VM Zavala (2014). ‘Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach’. In: *Energy and Buildings* 75, pages 189–198.

Carlos, JS and MC Nepomuceno (2012). ‘A simple methodology to predict heating load at an early design stage of dwellings’. In: *Energy and Buildings* 55, pages 198–207.

Carlucci, S and L Pagliano (2012). ‘A review of indices for the long-term evaluation of the general thermal comfort conditions in buildings’. In: *Energy and Buildings* 53, pages 194–205.

Caruso, G and JH Kämpf (2015). ‘Building shape optimisation to reduce air-conditioning needs using constrained evolutionary algorithms’. In: *Solar Energy* 118, pages 186–196.

Cebecauer, T and M Suri (2015). ‘Typical Meteorological Year Data: SolarGIS Approach’. In: *Energy Procedia*. International Conference on Concentrating Solar Power and Chemical Energy Systems, SolarPACES 2014 69, pages 1958–1969.

Chakraborty, D, H Elzarka and R Bhatnagar (2016). 'Generation of accurate weather files using a hybrid machine learning methodology for design and analysis of sustainable and resilient buildings (accepted manuscript)'. In: *Sustainable Cities and Society*.

Chan, ALS (2016). 'Generation of typical meteorological years using genetic algorithm for different energy systems'. In: *Renewable Energy* 90, pages 1–13.

Chinazzo, G (2014). 'Refurbishment of Existing Envelopes in Residential Buildings: assessing robust solutions for future climate change'. MSc. Lausanne, Switzerland: EPFL.

Chinazzo, G, P Rastogi and M Andersen (2015a). 'Assessing robustness regarding weather uncertainties for energy-efficiency-driven building refurbishments'. In: *Proceedings of IBPC 2015*. Torino.

– (2015b). 'Robustness Assessment Methodology for the Evaluation of Building Performance With a View to Climate Uncertainties'. In: *Proceedings of BS 2015*. Hyderabad, India.

Chow, TT, A Chan, KF Fong and Z Lin (2006). 'Some perceptions on typical weather year - from the observations of Hong Kong and Macau'. In: *Solar Energy* 80.4, pages 459–467.

Christensen, R (1991). *Linear models for multivariate, time series, and spatial data*. Springer texts in statistics. New York: Springer-Verlag.

CIBSE (2005). *Climate change and the indoor environment: impacts and adaptation*. English. TM36. London.

Clarke, JA (2001). *Energy Simulation in Building Design*. 2nd edition. Butterworth-Heinemann.

– (2015). 'A vision for building performance simulation: a position paper prepared on behalf of the IBPSA Board'. en. In: *Journal of Building Performance Simulation* 8.2, pages 39–43.

Coffey, B (2012). 'Using Building Simulation and Optimization to Calculate Lookup Tables for Control'. In:

Coley, D and T Kershaw (2010). 'Changes in internal temperatures within the built environment as a response to a changing climate'. In: *Building and Environment*. International Symposium on the Interaction between Human and Building Environment Special Issue Section 45.1, pages 89–93.

Collins, L, S Natarajan and GJ Levermore (2010). 'Climate change and future energy consumption in UK housing stock'. en. In: *Building Services Engineering Research and Technology* 31.1, pages 75–90.

Cox, RA, M Drews, C Rode and SB Nielsen (2015). 'Simple future weather files for estimating heating and cooling demand'. In: *Building and Environment*. Special Issue: Climate adaptation in cities 83, pages 104–114.

Crawley, DB (2007). 'Creating weather files for climate change and urbanization impacts analysis'. In: *Proceedings of BS 2007*. Volume 7. Beijing, China, pages 1075–1082.

– (2008). 'Estimating the impacts of climate change and urbanization on building performance'. In: *Journal of Building Performance Simulation* 1.2, pages 91–115.

– (2011). *Weather Data in Building Design and Performance Simulation*. Conference. Boston, MA.

Crawley, DB, JW Hand, M Kummert and BT Griffith (2005). *Contrasting the Capabilities of Building Energy Performance Simulation Programs*. Technical report v 1.0.

– (2008). 'Contrasting the capabilities of building energy performance simulation programs'. In: *Building and environment* 43.4, pages 661–673.

Crawley, DB and YJ Huang (1997). 'Does it matter which weather data you use in energy simulations'. In: *User News* 18.1, pages 25–31.

Crawley, DB and LK Lawrie (2015a). *Climate.OneBuilding.Org*.

– (2015b). 'Rethinking the TMY: Is the 'Typical' Meteorological Year Best for Building Performance Simulation?' In: *Proceedings of BS 2015*. Hyderabad, India.

Crow, LW (1981). *Development of Hourly Data for Weather Year for Energy Calculations (WYEC), Including Solar Data, at 21 Stations Throughout the U.S. | ASHRAE Store*. Technical report CH-81-12-2 (RP-239), pages 896–905.

Cryer, JD and KS Chan (2008). *Time Series Analysis: With Applications in R*. en. Springer.

David, M, L Adelard, F Garde and H Boyer (2005). 'Weather data analysis based on typical weather sequence analysis. Application: energy building simulation'. In: *Proceedings of BS 2005*. Montréal: IBPSA.

David, M, L Adelard, P Lauret and F Garde (2010). 'A method to generate Typical Meteorological Years from raw hourly climatic databases'. In: *Building and Environment* 45.7, pages 1722–1732.

Davison, AC (2003). *Statistical Models*. en. Cambridge University Press.

– (2013). 'Time Series'. English. Lausanne, Switzerland.

Davison, AC and DV Hinkley (1997). *Bootstrap Methods and their Application*. English. 1st. Cambridge University Press.

Davison, AC and CL Tsai (1992). 'Regression Model Diagnostics'. In: *International Statistical Review / Revue Internationale de Statistique* 60.3, pages 337–353.

De Dear, RJ (2006). 'Adapting buildings to a changing climate: but what about the occupants?' en. In: *Building Research & Information* 34.1, pages 78–81.

De Dear, RJ and GS Brager (1998). 'Developing an adaptive model of thermal comfort and preference'. In: *ASHRAE Transactions* 104.1, pages 145–167.

– (2002). 'Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55'. In: *Energy and Buildings* 34.6, pages 549–561.

De Livera, AM, RJ Hyndman and RD Snyder (2011). 'Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing'. In: *Journal of the American Statistical Association* 106.496, pages 1513–1527.

De Souza, CB (2012). 'Contrasting paradigms of design thinking: The building thermal simulation tool user vs. the building designer'. In: *Automation in Construction* 22, pages 112–122.

– (2013). 'Studies into the use of building thermal physics to inform design decision making'. In: *Automation in Construction* 30, pages 81–93.

De Souza, CB and I Knight (2007). 'Thermal Performance Simulation from an Architectural Design Viewpoint'. In: *Proceedings of BS 2007*. Beijing, China: IBPSA, pages 87–94.

De Wilde, P and D Coley (2012). 'The implications of a changing climate for buildings'. In: *Building and Environment* 55, pages 1–7.

De Wilde, P, Y Rafiq and M Beck (2008). 'Uncertainties in predicting the impact of climate change on thermal performance of domestic buildings in the UK'. In: *Building Service Engineering Research and Technology* 29.1, pages 7–26.

De Wilde, P and W Tian (2010). 'Predicting the performance of an office under climate change: A study of metrics, sensitivity and zonal resolution'. In: *Energy and Buildings* 42.10, pages 1674–1684.

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

– (2003). 'Uncertainty in Building Simulation'. In: *Advanced building simulation*. Edited by A Malkawi and G Augenbroe. New York: Spon Press.

Degelman, LO (1976). 'A weather simulation model for building energy analysis'. In: *ASHRAE Transactions* 82, pages 435–447.

– (1991). 'A statistically-based hourly weather data generator for driving energy simulation and equipment design software for buildings'. In: *Proceedings of Building Simulation*. Volume 91, pages 592–599.

– (1997). 'Examination of the Concept of Using "Typical-Week" Weather Data for Simulation of Annualized Energy Use in Buildings'. In: *Proceedings of BS 1997*. Prague: IBPSA.

– (2003). 'Simulation and Uncertainty: Weather Predictions'. In: *Advanced building simulation*. Edited by A Malkawi and G Augenbroe. New York: Spon Press.

Deru, M, K Field, D Studer, K Benne, B Griffith, P Torcellini, B Liu, M Halverson, D Winiarski, M Rosenberg, M Yazdanian, YJ Huang and DB Crawley (2011). *U.S. Department of Energy commercial reference building models of the national building stock*. Technical report, pages 1–118.

Dodge, Y (2008). *The concise encyclopedia of statistics*. 1st. ed. Springer reference. New York: Springer.

Donn, M (2009). *Simulation of Building Performance: Environmental Design Decision Support Tools in Architecture*. VDM Publishing.

Donn, M, S Selkowitz and B Bordass (2012). 'The building performance sketch'. In: *Building Research & Information* 40.2, pages 37–41.

Duvenaud, D (2014). 'Automatic model construction with Gaussian processes'. PhD thesis. University of Cambridge.

Duvenaud, D, JR Lloyd, R Grosse, JB Tenenbaum and Z Ghahramani (2013). 'Structure Discovery in Nonparametric Regression through Compositional Kernel Search'. In: *arXiv:1302.4922 [cs, stat]*. arXiv: 1302.4922.

Eames, M, T Kershaw and D Coley (2011). 'On the creation of future probabilistic design weather years from UKCP09'. en. In: *Building Services Engineering Research and Technology* 32.2, pages 127–142.

– (2012a). 'A comparison of future weather created from morphed observed weather and created by a weather generator'. In: *Building and Environment* 56, pages 252–264.

– (2012b). 'The appropriate spatial resolution of future weather files for building simulation'. In: *Journal of Building Performance Simulation* 5.6, pages 347–358.

Eames, M, M Wood and P Challenor (2015). 'The Implications of Transporting Architecture on Human Health'. In: *Proceedings of BS 2015*. Hyderabad, India: IBPSA.

Ebden, M (2008). *Gaussian Processes for Regression: A Quick Introduction*. Technical report. University of Oxford.

Eisenhower, B, Z O'Neill, S Narayanan, VA Fonoberov and I Mezić (2012). 'A methodology for meta-model based optimization in building energy models'. In: *Energy and Buildings* 47, pages 292–301.

Ellis, P, B Griffith and N Long (2006). 'Automated multivariate optimization tool for energy analysis'. In: *Proceedings of SB 2006*. MA, USA: IBPSA-USA.

Environment Canada (2015). *Canadian Weather year for Energy Calculation*. eng.

ESRU (2015). *ESP-r*.

Essenwanger, OM (2001). *General Climatology 1C: Classification of Climates*. en. Edited by HE Landsberg. Volume 1C. World Survey of Climatology. Elsevier Science Limited.

European Union (2010). *Directive 2010/31/EU of the European Parliament and of the Council, of 19 May 2010, on the energy performance of buildings (recast)*. EN.

Evins, R (2013). 'A review of computational optimisation methods applied to sustainable building design'. In: *Renewable and Sustainable Energy Reviews* 22, pages 230–245.

Fanger, PO (1970). *Thermal comfort: Analysis and applications in environmental engineering*. Copenhagen: Danish Technical Press.

Fathy, H, W Shearer and AR Sultan (1986). *Natural Energy and Venacular Architecture: Principles and Examples With Reference to Hot Arid Climates*. United Nations University.

Festa, R and C Ratto (1993). 'Proposal of a numerical procedure to select Reference Years'. In: *Solar Energy* 50.1, pages 9–17.

Finkelstein, J and R Schafer (1971). 'Improved goodness-of-fit tests'. In: *Biometrika* 58.3, pages 641–645.

Firth, SK, KJ Lomas and AJ Wright (2010). 'Targeting household energy-efficiency measures using sensitivity analysis'. In: *Building Research & Information* 38.1, pages 25–41.

Foucquier, A, S Robert, F Suard, L Stéphan and A Jay (2013). 'State of the art in building modelling and energy performances prediction: A review'. In: *Renewable and Sustainable Energy Reviews* 23, pages 272–288.

Fourier, JBJ (1822). *Théorie analytique de la chaleur*. fr. Chez Firmin Didot, püre et fils.

Frank, T (2005). ‘Climate change impacts on building heating and cooling energy demand in Switzerland’. In: *Energy and Buildings*. Research That Inspires 125 Years of EMPA 37.11, pages 1175–1185.

Fürbringer, JM and CA Roulet (1995). ‘Comparison and combination of factorial and Monte-Carlo design in sensitivity analysis’. In: *Building and Environment* 30.4.

Gagne, JML (2011). ‘An Interactive Performance-Based Expert System for Daylighting in Architectural Design’. Docto.

Gagne, JML and M Andersen (2010). ‘Multi-objective facade optimization for daylighting design using a genetic algorithm’. In: 9.

Gagne, JML, M Andersen and LK Norford (2011). ‘An interactive expert system for daylighting design exploration’. In: *Building and Environment* 46.11, pages 2351–2364.

Garcia Sanchez, D, B Lacarrière, M Musy and B Bourges (2014). ‘Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods’. In: *Energy and Buildings* 68, Part C, pages 741–750.

Gaterell, M and M McEvoy (2005). ‘The impact of climate change uncertainties on the performance of energy efficiency measures applied to dwellings’. In: *Energy and Buildings* 37.9, pages 982–995.

Gazela, M and E Mathioulakis (2001). ‘A new method for typical weather data selection to evaluate long-term performance of solar energy systems’. In: *Solar Energy* 70.4, pages 339–348.

German Weather Service (DWD) (2010). *Updated and enhanced test reference years (TRY) of Germany for medium and extreme weather conditions*.

Gervásio, H, P Santos, R Martins and L Simões da Silva (2014). ‘A macro-component approach for the assessment of building sustainability in early stages of design’. In: *Building and Environment* 73, pages 256–270.

Givoni, B (1976). *Man, climate and architecture*. Applied Science Publishers.

– (1989). *Urban Design in Different Climates*. Volume 10.

– (1992). ‘Comfort, climate analysis and building design guidelines’. In: *Energy and Buildings* 18.1, pages 11–23.

Glassman, EJ and C Reinhart (2013). ‘Façade optimization using parametric design and future climate scenarios’. In: *Proceedings of BS 2013*. Chambéry, France.

Granadeiro, V, JR Correia, VM Leal and JP Duarte (2013). ‘Envelope-related energy demand: A design indicator of energy performance for residential buildings in early design stages’. In: *Energy and Buildings* 61, pages 215–223.

Grondzik, WT, AG Kwok, B Stein and JS Reynolds (2011). *Mechanical and Electrical Equipment for Buildings*. en. John Wiley & Sons.

Guan, L (2009). ‘Preparation of future weather data to study the impact of climate change on buildings’. In: *Building and Environment* 44.4, pages 793–800.

Guan, L (2012). ‘Energy use, indoor temperature and possible adaptation strategies for air-conditioned office buildings in face of global warming’. In: *Building and Environment* 55, pages 8–19.

Guan, LS, J Yang and J Bell (2005). ‘A Method of Preparing Future Hourly Weather Data for the Study of Global Warming Impact on the Built Environment’. In: *Faculty of Built Environment and Engineering*. CD Rom: Queensland University of Technology, pages 1–12.

Gupta, R and D Dantsiou (2013). ‘Understanding the Gap between ‘as Designed’ and ‘as Built’ Performance of a New Low Carbon Housing Development in UK’. en. In: *Sustainability in Energy and Buildings*. Edited by A Hakansson, M Höjer, RJ Howlett and LC Jain. Smart Innovation, Systems and Technologies 22. DOI: 10.1007/978-3-642-36645-1_53. Springer Berlin Heidelberg, pages 567–580.

Haghighat, F and A Athienitis (1988). ‘Comparison between time domain and frequency domain computer program for building energy analysis’. In: *Computer-Aided Design* 20.9, pages 525–532.

Haghighat, F, M Chandrashekar and T Unny (1987). ‘Thermal behaviour of buildings under random conditions’. In: *Applied Mathematical Modelling* 11.5, pages 349–356.

Haghighat, F, TE Unny and M Chandrashekar (1985). ‘Stochastic Modeling of Transient Heat Flow Through Walls’. In: *Journal of Solar Energy Engineering* 107.3, pages 202–207.

Halawa, E and J van Hoof (2012). ‘The adaptive approach to thermal comfort: A critical overview’. In: *Energy and Buildings* 51, pages 101–110.

Haldi, F (2010). ‘Towards a Unified Model of Occupants’ Behaviour and Comfort for Building Energy Simulation’. Doctoral. Ecole polytechnique fédérale de Lausanne.

– (2013). ‘A Probabilistic Model to Predict Building Occupants’ Diversity towards their Interactions with the Building Envelope’. In: *Proceedings of BS 2013*. Chambéry, France.

Hall, I, R Prairie, H Anderson and E Boes (1978). ‘Generation of a typical meteorological year’. In:

Hallgreen, L (1983). ‘Short Reference Year, SRY’. en. In: *Solar Radiation Data*. Edited by W Palz. Volume 2. Solar Energy R&D in the European Community Series F. DOI: 10.1007/978-94-009-7112-7_5. Springer Netherlands, pages 40–48.

Hamby, DM (1994). ‘A review of techniques for parameter sensitivity analysis of environmental models’. en. In: *Environmental Monitoring and Assessment* 32.2, pages 135–154.

Hansen, JE and DM Driscoll (1977). ‘A Mathematical Model for the Generation of Hourly Temperatures’. In: *Journal of Applied Meteorology* 16.9, pages 935–948.

Harriman, LG (2008). *The ASHRAE Guide for Buildings in Hot and Humid Climates*. American Society of Heating, Refrigerating & Air-Conditioning Engineers, Incorporated.

Harriman, LG, DG Colliver and KQ Hart (1999). ‘New Weather Data For Energy Calculations’. en. In: *ASHRAE Journal* 41.3, pages 31–37.

Hausladen, G, M de Saldanha and P Liedl (2012). *Building to suit the climate: A Handbook*. Basel: Birkhäuser.

Helton, J, J Johnson, C Sallaberry and C Storlie (2006). ‘Survey of sampling-based methods for uncertainty and sensitivity analysis’. en. In: *Reliability Engineering & System Safety* 91.10-11, pages 1175–1209.

Hensen, J (1990). ‘Literature review on thermal comfort in transient conditions’. In: *Building and Environment* 25.4, pages 309–316.

Heo, Y and VM Zavala (2012). ‘Gaussian process modeling for measurement and verification of building energy savings’. In: *Energy and Buildings* 53, pages 7–18.

Heschong, L (1979). *Thermal Delight in Architecture*. en. MIT Press.

Hitchcock, RJ, ES Lee and C Huizenga (2008). ‘COMFEN: A Commercial Fenestration/ Façade Design Tool’. In: pages 246–252.

Hitchin, ER, MJ Holmes, BC Hutt, S Irving and D Nevrala (1983). ‘The CIBS example weather year’. en. In: *Building Services Engineering Research and Technology* 4.3, pages 119–124.

Hodge, BK (2010). *Alternative energy systems and applications*. Hoboken, NJ: Wiley.

Hoefer, C (2015). *Causal Determinism*. Edited by EN Zalta.

Hoes, P, M Trcka, J Hensen and B Bonnema (2011). ‘Optimizing building designs using a robustness indicator with respect to user behavior’. In: *Proceedings of BS 2011*, pages 1710–1717.

Holman, JP (1972). *Heat transfer*. 3rd edition. McGraw-Hill.

Holmes, MJ and ER Hitchin (1978). ‘An Example Year for the Calculation of Energy Demand in Buildings’. In: *Building Service Engineering Research and Technology* 45, pages 186–189.

Hong, T, WK Chang and HW Lin (2013). ‘A sensitivity study of building performance using 30-year actual weather data’. In: *Proceedings of BS 2013*. Chambéry, France.

Hong, T and Y Jiang (1995). ‘Stochastic Weather Model for Building HVAC Systems’. In: *Building and Environment* 30.4, pages 521–532.

Hoof, Jv (2010). ‘Thermal comfort: research and practice’. In: *Frontiers in Bioscience* 15.1, page 765.

Hopfe, CJ, GL Augenbroe and JL Hensen (2013). ‘Multi-criteria decision making under uncertainty in building performance assessment’. In: *Building and Environment* 69, pages 81–90.

Hopfe, CJ and JLM Hensen (2011). ‘Uncertainty analysis in building performance simulation for design support’. In: *Energy and Buildings* 43.10, pages 2798–2805.

Hopfe, CJ (2009). ‘Uncertainty and sensitivity analysis in building performance simulation for decision support and design optimization’. en. PhD. Eindhoven, The Netherlands: Technische Universiteit Eindhoven.

Hotelling, H (1933). ‘Analysis of a complex of statistical variables into principal components.’ In: *Journal of educational psychology* 24.6, page 417.

Huang, KT and RL Hwang (2015). ‘Future trends of residential building cooling energy and passive adaptation measures to counteract climate change: The case of Taiwan (in press)’. In: *Applied Energy*.

Huang, YJ (2012). *International Weather for Energy Calculations*. ASHRAE Research Project RP-1477. Atlanta, GA, USA: ASHRAE.

– (2015). *White Box Technologies*.

Hubbard, KG, KE Kunkel, KT Redmond and AT Degaetano (2005). 'Sources of uncertainty in the calculation of design weather conditions'. In: volume 111, pages 317–326.

Hui, Cm (1996). 'Energy performance of air-conditioned buildings in Hong Kong'. PhD. City University of Hong Kong.

Hui, SC and K Cheung (1997). 'Multi-year (MY) building simulation: is it useful and practical?' In: *Proc. of the IBPSA Building Simulation'97 Conference*, pages 8–10.

Hulme, M, editor (2002). *Climate change scenarios for the United Kingdom: the UKCIP02 scientific report*. Norwich, UK: Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia.

Humphreys, M (1978). 'Outdoor temperatures and comfort indoors'. In: *Batiment International, Building Research and Practice* 6.2, pages 92–92.

Humphreys, MA, F Nicol and S Roaf (2016). *Adaptive thermal comfort: foundations and analysis*. London ; New York: Routledge, Taylor & Francis Group.

Humphreys, M, F Nicol, O Sykes and S Roaf, editors (1995). *Standards for Thermal Comfort: Indoor air temperature standards for the 21st century*. en. 1st. Taylor & Francis.

Hygh, JS, JF DeCarolis, DB Hill and SR Ranjithan (2012). 'Multivariate regression as an energy assessment tool in early building design'. In: *Building and Environment* 57, pages 165–175.

Iaccarino, G (2008). *Quantification of Uncertainty in Flow Simulations Using Probabilistic Methods*. Hosted by the website of the Uncertainty Quantification Laboratory at Stanford University. Stanford University.

IBPSA-USA (2015). *BEST Directory: Building Energy Software Tools*.

Indraganti, M, R Ooka, HB Rijal and GS Brager (2014). 'Adaptive model of thermal comfort for offices in hot and humid climates of India'. In: *Building and Environment* 74, pages 39–53.

International Code Council (2012). *International Energy Conservation Code*. English.

IPCC (2013). *Stochastic weather generators*.

– (2014a). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Technical report. Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC).

– (2014b). *Climate Change 2014 Synthesis Report: Summary for Policymakers*. Technical report. Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC).

Jain, N, AP Ramallo-González and S Natarajan (2014). 'Effects of Aggressive Energy Efficiency Regulations on an Unprepared Building Sector using Uncertainty Analysis'. In: *Proceedings of PLEA 2014*. Ahmedabad: University of Bath.

Janda, KB (2011). 'Buildings don't use energy: people do'. en. In: *Architectural Science Review* 54.1, pages 15–22.

Jenkins, DP, S Patidar, PFG Banfill and GJ Gibson (2011). 'Probabilistic climate projections with dynamic building simulation: Predicting overheating in dwellings'. In: *Energy and Buildings* 43.7, pages 1723–1731.

Jenkins, DP, S Patidar and SA Simpson (2015). 'Quantifying Change in Buildings in a Future Climate and Their Effect on Energy Systems'. en. In: *Buildings* 5.3, pages 985–1002.

Jenkins, D, M Gul and S Patidar (2013). 'Probabilistic future cooling loads for mechanically cooled offices'. In: *Energy and Buildings* 66, pages 57–65.

Jentsch, MF, AS Bahaj and PAB James (2008). 'Climate change future proofing of buildings - Generation and assessment of building simulation weather files'. In: *Energy and Buildings* 40.12, pages 2148–2168.

Jentsch, MF, M Eames and GJ Levermore (2015). 'Generating near-extreme Summer Reference Years for building performance simulation'. en. In: *Building Services Engineering Research and Technology*, page 0143624415587476.

Jentsch, MF, PA James, L Bourikas and AS Bahaj (2013). 'Transforming existing weather data for worldwide locations to enable energy and building performance simulation under future climates'. In: *Renewable Energy* 55, pages 514–524.

Jin, Q and M Overend (2014). 'Sensitivity of façade performance on early-stage design variables'. In: *Energy and Buildings* 77, pages 457–466.

Jones, P, C Harpham, V Glenis and A Burton (2010). *UK Climate Projections science report: Projections of future daily climate for the UK from the Weather Generator (revised November 2010)*. Technical report. Exeter: Met Office Hadley Centre.

Judkoff, R and J Neymark (1995). *International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method*. Technical report. NREL and IEA.

– (2006). 'Model validation and testing: the methodological foundation of ASHRAE standard 140'. In: *Transactions-American Society of Heating Refrigerating and Air Conditioning Engineers* 112.2, page 367.

Kalogirou, SA (2003). 'Generation of typical meteorological year (TMY-2) for Nicosia, Cyprus'. In: *Renewable Energy* 28.15, pages 2317–2334.

Kershaw, T, M Eames and D Coley (2010). 'Comparison of multi-year and reference year building simulations'. en. In: *Building Services Engineering Research and Technology* 31.4, pages 357–369.

– (2011). 'Assessing the risk of climate change for buildings: A comparison between multi-year and probabilistic reference year simulations'. In: *Building and Environment* 46.6, pages 1303–1308.

Kim, SH and G Augenbroe (2013). 'Uncertainty in developing supervisory demand-side controls in buildings: A framework and guidance'. In: *Automation in Construction* 35, pages 28–43.

Kim, YJ, KU Ahn, C Park and IH Kim (2013). 'Gaussian emulator for stochastic optimal design of a double glazing system'. In: *Proceedings of the 13th IBPSA Conference, August*, pages 25–28.

Kleijnen, JPC (2001). 'Experimental Designs for Sensitivity Analysis of Simulation Models'. In: edited by A Heemink. Volume 2001. Delft University of Technology, pages 26–29.

– (2008). *Design and Analysis of Simulation Experiments*. Volume 111. International Series in Operations Research & Management Science. Boston, MA: Springer US.

– (2009). 'Kriging metamodeling in simulation: A review'. In: *European Journal of Operational Research* 192.3, pages 707–716.

Kleijnen, JPC and Wv Groenendaal (1992). *Simulation: a statistical perspective*. eng. Chichester, West Sussex, England ; New York: J. Wiley.

Koenigsberger, OH, TG Ingersoll, A Mayhew and SV Szokolay (1974). *Manual of Tropical Housing and Building: Climatic design*. en. Longman.

Kottek, M, J Grieser, C Beck, B Rudolf and F Rubel (2006). 'World Map of the Koeppen-Geiger climate classification updated'. In: *Meteorologische Zeitschrift* 15.3, pages 259–263.

Krarti, M (2000). *Energy audit of building systems: an engineering approach*. Mechanical engineering series. Boca Raton, FL: CRC Press.

Kuhn, TS (2012). *The structure of scientific revolutions*. 4th ed. Chicago; London: The University of Chicago Press.

Kwong, QJ, NM Adam and B Sahari (2013). 'Thermal Comfort Assessment and Potential for Energy Efficiency Enhancement in Modern Tropical Buildings: A Review'. In: *Energy and Buildings*.

Lall, U and A Sharma (1996). 'A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series'. In: *Water Resources Research* 32.3, pages 679–679.

Lam, JC, CL Tsang, L Yang and DHW Li (2005). 'Weather data analysis and design implications for different climatic zones in China'. In: *Building and Environment* 40.2, pages 277–296.

Lam, KP, YC Huang and C Zhai (2004). *Energy Modeling Tools Assessment For Early Design Phase*. Technical report. Pittsburgh, PA, USA, pages 79–79.

Law, T (2013). 'The Future of Thermal Comfort in an Energy-Constrained World'. DOI: 10.1007/978-3-319-00149-4. PhD thesis. Launceston, Australia: University of Tasmania.

Lazos, D, AB Sproul and M Kay (2015). 'Development of hybrid numerical and statistical short term horizon weather prediction models for building energy management optimisation'. In: *Building and Environment* 90, pages 82–95.

Lee, BD, Y Sun, G Augenbroe and CJ Paredis (2013). 'Towards better prediction of building performance: A workbench to analyze uncertainty in building simulation'. In: *Proceedings of BS 2013*. Chambéry, France: IBPSA.

Lee, BD, Y Sun, H Hu, G Augenbroe and CJJ Paredis (2012). 'A Framework For Generating Stochastic Meteorological Years For Risk-Conscious Design Of Buildings'. In: pages 345–352.

Levermore, GJ, R Courtney, R Watkins, H Cheung, JB Parkinson, P Laycock, S Natarajan, M Nikolopoulou, C McGilligan, T Muneer, Y Tham, CP Underwood, JS Edge, H Du, S Sharples, J Kang, M Barclay and M Sanderson (2012). *Deriving and using future weather data for* *building design from UK climate change projections – an overview of the COPSE Project*. United Kingdom: University of Manchester.

Levermore, GJ and JB Parkinson (2006). 'Analyses and algorithms for new Test Reference Years and Design Summer Years for the UK'. In: *Building Service Engineering Research and Technology* 27.4, pages 311–325.

Lindelöf, D (2007). 'Bayesian optimization of visual comfort'. eng. Doctoral. Lausanne: Ecole polytechnique fédérale de Lausanne.

Lomas, KJ and H Eppel (1992). 'Sensitivity analysis techniques for building thermal simulation programs'. In: *Energy and Buildings* 19.1, pages 21–44.

Loonen, R and JLM Hensen (2013). 'Dynamic sensitivity analysis for performance-based building design and operation'. In: *Proceedings of BS 2013*. Chambéry, France: IBPSA.

Lund, H (1991). 'The Design Reference Year'. In: *Proceedings of Building Simulation 1991*. Nice, France: IBPSA, pages 600–606.

– (1995). *The Design Reference Year Users Manual - A Report of Task 9 : Solar Radiation and Pyranometer Studies*. Technical report. Copenhagen, Denmark.

– (2001). 'Design Reference Years and Test Reference Years in Europe, Turkey and Israel'. In: 262.

Macdonald, I, J Clarke and P Strachan (1999). 'Assessing uncertainty in building simulation'. In: *Proc. Building Simulation*. Volume 99.

Macdonald, I (2002). 'Quantifying the Effects of Uncertainty in Building Simulation'. Doctoral. University of Strathclyde.

Macdonald, I and P Strachan (2001). 'Practical application of uncertainty analysis'. In: *Energy and Buildings* 33.3, pages 219–227.

Mackay, DJ (1998). *Introduction to Gaussian Processes*. Technical report. Cambridge, UK: Cambridge University.

Madeddu, D (2011). 'Interactive facade optimized for daylighting and pedestrian response using a genetic algorithm.' In: pages 508–514.

Magnano, L, J Boland and RJ Hyndman (2008). 'Generation of synthetic sequences of half-hourly temperature'. en. In: *Environmetrics* 19.8, pages 818–835.

Magnano, L (2007). 'Mathematical models for temperature and electricity demand'. In:

Mahdavi, A and KP Lam (1993). 'A Dialectic of Process and Tool: Knowledge Transfer and Decision-Making Strategies in the Building Delivery Process'. In: edited by KS Mathur, MP Betts and KW Tham. World Scientific Pub Co Inc, pages 345–356.

Mahdavi, A and P Mahattanatawe (2003). 'Enclosure systems design and control support via dynamic simulation-assisted optimization'. In: *Proceedings of BS 2003*. Eindhoven, Netherlands: IBPSA, pages 785–792.

Manley, G (1974). 'Central England temperatures: monthly means 1659 to 1973'. In: *Quarterly Journal of the Royal Meteorological Society* 100.425, pages 389–405.

Manu, S, Y Shukla, R Rawal, LE Thomas and R de Dear (2016). 'Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC)'. In: *Building and Environment* 98, pages 55–70.

Marijt, R (2009). 'Multi-objective Robust Optimization Algorithms for Improving Energy Consumption and Thermal Comfort of Buildings'. PhD thesis. Technical University of Eindhoven.

Marion, W and K Urban (1995). *User's Manual for TMY2*. Technical report. Golden, CO, USA: NREL, National Renewable Energy Laboratory.

Markus, TA, JA Clarke, EN Morris and TG Collins (1984). 'The influence of climate on housing: A simple technique for the assessment of dynamic energy behaviour'. In: *Energy and Buildings* 7.3, pages 243–259.

Markus, T (1982). 'Development of a cold climate severity index'. In: *Energy and Buildings* 4, pages 277–283.

Marsh, A and D Carruthers (1995). 'A selection of Interactive Design Tools'. In: Canberra, Australia.

Mavrogianni, A, M Davies, J Taylor, Z Chalabi, P Biddulph, E Oikonomou, P Das and B Jones (2014). 'The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments'. In: *Building and Environment* 78, pages 183–198.

Mayer, J, K Khairy and J Howard (2010). 'Drawing an elephant with four complex parameters'. en. In: *American Journal of Physics* 78.6. See http://www.johndcook.com/blog/2011/06/21/how-to-fit-an-elephant/ for a Python implementation., page 648.

Mazo, J, AT El Badry, J Carreras, M Delgado, D Boer and B Zalba (2015). 'Uncertainty propagation and sensitivity analysis of thermo-physical properties of phase change materials (PCM) in the energy demand calculations of a test cell with passive latent thermal storage'. en. In: *Applied Thermal Engineering* 90, pages 596–608.

McCullagh, P and JA Nelder (1983). *Generalized Linear Models*. Monographs on statistics and applied probability. London ; New York: Chapman and Hall.

McGilligan, C, S Natarajan and M Nikolopoulou (2011). 'Adaptive Comfort Degree-Days: A metric to compare adaptive comfort standards and estimate changes in energy consumption for future UK climates'. In: *Energy and Buildings* 43.10, pages 2767–2778.

McKay, MD, RJ Beckman and WJ Conover (1979). 'A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code'. In: *Technometrics* 21.2, pages 239–245.

McKitrick, RR and PJ Michaels (2007). 'Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data'. In: *Journal of Geophysical Research* 112.D24, D24S09–D24S09.

McQuiston, FC, JD Parker and JD Spitler (2005). *Heating, ventilating, and air conditioning: analysis and design*. en. John Wiley & Sons.

Meteorological Service of Canada and National Research Council of Canada (2008). *Canadian Weather Energy and Engineering Data Sets (CWEEDS Files) and Canadian Weather for Energy Calculations (CWEC Files): Updated User's Manual.*

MeteoSwiss (2014). *IDAWEB.*

Miller, AA (1961). *Climatology.* Methuen & Co.

MINERGIE Building Agency (2008). *The MINERGIE Standard for Buildings.* Technical report. Bern, Switzerland: MINERGIE.

Morton, T and P Bretschneider (2011). 'Building a better future: An exploration of beliefs about climate change and perceived need for adaptation within the building industry'. In: *Building and Environment* 46.5, pages 1151–1158.

Murdoch, N and JM Penman (1991). 'Building energy estimation by fast simulation'. In: *Solar Energy* 47.6, pages 447–455.

Murray, SN, BP Walsh, D Kelliher and DTJ O'Sullivan (2014). 'Multi-variable optimization of thermal energy efficiency retrofitting of buildings using static modelling and genetic algorithms – A case study'. In: *Building and Environment* 75, pages 98–107.

Mylona, A (2012). 'The use of UKCP09 to produce weather files for building simulation'. en. In: *Building Services Engineering Research and Technology* 33.1, pages 51–62.

Naboni, E, A Maccarini, I Korolija and Y Zhang (2013). 'Comparison of conventional, parametric and evolutionary optimization approaches for the architectural design of nearly zero energy buildings'. eng. In: *Proceedings of BS 2013.* Chambéry, France, pages 2558–2565.

Naboni, E, Y Zhang, A Maccarini, E Hirsch and D Lezzi (2013). 'Extending the use of parametric simulation in practice through a cloud based online service'. In: *Proceedings of BSA 2013.* Bozen, Italy.

Natarajan, S and GJ Levermore (2007). 'Predicting future UK housing stock and carbon emissions'. In: *Energy Policy* 35.11, pages 5719–5727.

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

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

Nault, E, P Rastogi, E Rey and M Andersen (2015). 'The sensitivity of predicted energy use to urban geometrical factors in various climates'. In: *Proceedings of PLEA 2015.* Bologna, Italy.

NCDC/NOAA (2014). *NNDC Climatic Data OnLine.*

Nicol, F (2004). 'Adaptive thermal comfort standards in the hot–humid tropics'. In: *Energy and Buildings* 36.7, pages 628–637.

Nicol, F and M Humphreys (1973). 'Thermal comfort as part of a self-regulating system'. In: *Building Research and Practice* 1.3, pages 174–179.

Nicol, F and M Humphreys (2002). ‘Adaptive thermal comfort and sustainable thermal standards for buildings’. In: *Energy and Buildings* 34.6, pages 563–572.

– (2010). ‘Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251’. In: *Building and Environment* 45.1, pages 11–17.

Nicol, F, MA Humphreys and S Roaf (2012). *Adaptive thermal comfort: principles and practice.* London ; New York: Routledge.

Nicol, F and M Wilson (2010). ‘An overview of the European Standard EN 15251’. In: *Adapting to Change: New Thinking on ...* April, pages 9–11.

Nik, VM, É Mata and A Sasic Kalagasidis (2015). ‘A statistical method for assessing retrofitting measures of buildings and ranking their robustness against climate change’. In: *Energy and Buildings* 88, pages 262–275.

NIST (2013). *NIST/SEMATECH e-Handbook of Statistical Methods.*

NREL and USDOE (2015). *EnergyPlus.*

Ochoa, CE and IG Capeluto (2009). ‘Advice tool for early design stages of intelligent facades based on energy and visual comfort approach’. In: *Energy and Buildings* 41.5, pages 480–488.

Of Architects, TAI (2012). *An Architect’s Guide to integrating energy modeling in the design process.*

Oko, COC and OB Ogoloma (2011). ‘Generation of a typical meteorological year for Port Harcourt zone’. In: *Journal of Engineering Science and Technology* 6.2, pages 204–214.

Olgyay, V and A Olgyay (1992). *Design with climate : bioclimatic approach to architectural regionalism.* Van Nostrand Reinhold.

Oliver, JE (2005). *Encyclopedia of world climatology.* Springer.

Oraiopoulos, A, T Kane, SK Firth and KJ Lomas (2015a). ‘Measured Internal Temperatures in UK Homes - A Time Series Analysis and Modelling Approach’. In: *Proceedings of BS 2015.* Hyderabad, India: IBPSA.

– (2015b). ‘Measured Internal Temperatures in UK Homes – A Time Series Analysis And Modelling Approach’. In: *Energy Procedia.* 6th International Building Physics Conference, IBPC 2015 78, pages 2844–2850.

Orlov, AI (2011). *Mahalanobis distance - Encyclopedia of Mathematics.*

Owen, MS, editor (2009). *ASHRAE Handbook: Fundamentals.* Atlanta, GA: American Society of Heating Refrigerating and Air-Conditioning Engineers.

– editor (2013). *ASHRAE Handbook: Fundamentals.* Atlanta, GA: American Society of Heating Refrigerating and Air-Conditioning Engineers.

Oxford English Dictionary (2016). *"synoptic, adj. (and n.)"*

Pacific Northwest National Laboratory and Oak Ridge National Laboratory (2010). *High-Performance Home Technologies: Guide to Determining Climate Regions by County.* Technical report PNNL-17211.

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

Pastore, L, P Rastogi, S Rockcastle, G Rueff, H Monari and M Andersen (2016). 'Assessing the impact of contemporary urbanization on bioclimatic features of historic architecture through a two-step simulation process'. In: Los Angeles, USA.

Patidar, S, DP Jenkins, PFG Banfill and GJ Gibson (2012). 'Simple statistical model for complex probabilistic climate projections: Overheating risk and extreme events'. In: *Renewable Energy*, pages 1–6.

Patidar, S, DP Jenkins, GJ Gibson and PFG Banfill (2011). 'Statistical techniques to emulate dynamic building simulations for overheating analyses in future probabilistic climates'. In: *Journal of Building Performance Simulation* 4.3, pages 271–284.

– (2012). 'Correlating Probabilistic Climate Projections With Cooling Demand In An Office Building'. In: IBPSA-England, pages 261–268.

Pearson, K (1901). 'On lines and planes of closest fit to systems of points in space'. In: *Philosophical Magazine* 2.11, pages 559–572.

Peel, MC, BL Finlayson and TA Mcmahon (2007). 'Updated world map of the Koeppen-Geiger climate classification'. In: *Hydrology and Earth System Sciences* 11.5, pages 1633–1644.

Perera, DWU, M Halstensen and NO Skeie (2015). 'Prediction of Space Heating Energy Consumption in Cabins Based on Multivariate Regression Modelling'. In: *International Journal of Modeling and Optimization* 5.6, pages 385–392.

Perez, R, P Ineichen, R Seals, J Michalsky and R Stewart (1990). 'Modeling daylight availability and irradiance components from direct and global irradiance'. In: *Solar Energy* 44.5, pages 271–289.

Perez, R, R Seals, P Ineichen, R Stewart and D Menicucci (1987). 'A new simplified version of the perez diffuse irradiance model for tilted surfaces'. In: *Solar Energy* 39.3, pages 221–231.

Pianka, ER (2008). *Convergent Evolution.*

Politis, DN, JP Romano and M Wolf (1999). *Subsampling.* Springer series in statistics. New York: Springer.

Politis, D (1998). 'Computer-intensive methods in statistical analysis'. In: *IEEE Signal Processing Magazine* 15.1, pages 39–55.

Pont, U, F Shayeganfar, N Ghiassi, M Taheri, C Sustr, A Mahdavi, S Fenz, J Heurix, A Anjomshoaa, T Neubauer and AM Tjoa (2013). 'Recent advances in SEMERGY: A semantically enriched optimization environment for performance-guided building design and refurbishment'. In: *Contributions to Building Physics.* Edited by A Mahdavi and B Martens. Vienna, Austria.

Pranovich, S, JJ van Wijk and H van de Wetering (2003). 'An architectural design system for the early stages'. In:

Pyke, CR, S McMahon, L Larsen, NB Rajkovich and A Rohloff (2012). 'Development and analysis of Climate Sensitivity and Climate Adaptation opportunities indices for buildings'. In: *Building and Environment*. Implications of a Changing Climate for Buildings 55, pages 141–149.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Racsko, P, L Szeidl and M Semenov (1991). 'A serial approach to local stochastic weather models'. In: *Ecological Modelling* 57.1, pages 27–41.

Ramallo-González, AP, TS Blight and DA Coley (2015). 'New optimisation methodology to uncover robust low energy designs that accounts for occupant behaviour or other unknowns'. In: *Journal of Building Engineering* 2, pages 59–68.

Ramallo-González, AP and D Coley (2014). 'Using self-adaptive optimisation methods to perform sequential optimisation for low-energy building design'. In: *Energy and Buildings* 81, pages 18–29.

Rannow, S, W Loibl, S Greiving, D Gruehn and BC Meyer (2010). 'Potential impacts of climate change in Germany—Identifying regional priorities for adaptation activities in spatial planning'. In: *Landscape and Urban Planning* 98.3-4, pages 160–171.

Rasmussen, CE and CKI Williams (2006). *Gaussian processes for machine learning*. Adaptive computation and machine learning. Cambridge, Mass: MIT Press.

Rastogi, P and M Andersen (2013). 'Generation of Weather Files Using Resampling Techniques: An Exploratory Study'. In: *Proceedings of BS 2013*. Chambéry, France, pages 1422–1429.

– (2015). 'Embedding Stochasticity in Building Simulation Through Synthetic Weather Files'. In: *Proceedings of BS 2015*. Hyderabad, India.

– (2016). 'Incorporating Climate Change Predictions in the Analysis of Weather-Based Uncertainty'. In: *Proceedings of SimBuild 2016*. Salt Lake City, UT, USA.

Rastogi, P, SF Horn and M Andersen (2013). 'Toward Assessing the Sensitivity of Buildings to Changes in Climate'. In: *Proceedings of PLEA 2013*. Munich, Germany.

Remund, J, S Mueller, S Kunz and C Schilter (2012a). *METEONORM Handbook Part I : Software*. Technical report.

– (2012b). *METEONORM Handbook Part II : Theory*. Technical report.

Ren, M, N Shankland, N Holden, K Hague and J Webster (2012). 'Climate Change Adaptation Study For The Cooperative Head Office, Manchester'. In: IBPSA-England, pages 269–276.

Richardson, CW (1981). 'Stochastic simulation of daily precipitation, temperature, and solar radiation'. en. In: *Water Resources Research* 17.1, pages 182–190.

Roaf, S, D Crichton and F Nicol (2009). *Adapting buildings and cities for climate change: a 21st century survival guide*. eng. 2nd ed. Amsterdam: Elsevier.

Saltelli, A, editor (2008). *Sensitivity analysis*. eng. Paperback ed. Wiley paperback series. Chichester: Wiley.

Sandberg, NH, I Sartori and H Brattebø (2014). 'Sensitivity analysis in long-term dynamic building stock modeling—Exploring the importance of uncertainty of input parameters in Norwegian segmented dwelling stock model'. In: *Energy and Buildings* 85, pages 136–144.

Santner, TJ, BJ Williams and W Notz (2003). *The Design and analysis of computer experiments*. Springer series in statistics. New York: Springer.

Santos, P, R Martins, H Gervásio and L Simões da Silva (2014). 'Assessment of building operational energy at early stages of design – A monthly quasi-steady-state approach'. In: *Energy and Buildings* 79, pages 58–73.

Scartezzini, JL, F Bottazzi and MN Ferguson (1987). *Applying stochastic methods to building thermal design and control*. Synthesis Report NEFF 349 I FN 2.331-0.86. Lausanne, Switzerland.

– (1989). *Applying stochastic methods to building thermal design and control*. Final Report NEFF 349 I FN 2.331-0.86. Lausanne, Switzerland.

Scartezzini, JL, M Nygard Ferguson and F Bochud (1990). *Compression of multi-year meteorological data*. Technical report EF-REN (90) 009. Lausanne, Switzerland: Ecole polytechnique fédérale de Lausanne.

Schlueter, A and F Thesseling (2009). 'Building information model based energy/exergy performance assessment in early design stages'. In: *Automation in Construction* 18.2, pages 153–163.

Schröder, J and K Gawron (1981). 'Latent heat storage'. en. In: *International Journal of Energy Research* 5.2, pages 103–109.

Scott, MJ, LE Wrench and DL Hadley (1994). 'Effects of Climate Change on Commercial Building Energy Demand'. In: *Energy Sources* 16.3, pages 317–332.

Sengupta, M (2014). *India Solar Resource Data*.

Shen, H and A Tzempelikos (2012). 'Sensitivity analysis on daylighting and energy performance of perimeter offices with automated shading'. In: *Building and Environment* 59, pages 303–314.

Shumway, RH and DS Stoffer (2011). *Time Series Analysis and Its Applications: With R Examples*. English. New York, NY: Springer.

Silva, AS and E Ghisi (2014a). 'Uncertainty analysis of the computer model in building performance simulation'. In: *Energy and Buildings* 76, pages 258–269.

– (2014b). 'Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation'. In: *Energy and Buildings* 76, pages 381–391.

'Simulation of building energy and indoor environmental quality - some weather data issues' (1999). In: Prague, Czech Republic: Czech Hydrometeorological Institute.

Skeiker, K (2009). 'Optimum tilt angle and orientation for solar collectors in Syria'. In: *Energy Conversion and Management* 50.9, pages 2439–2448.

Smith, G, J Aguilar, a Gentle and D Chen (2012). 'Multi-parameter sensitivity analysis: A design methodology applied to energy efficiency in temperate climate houses'. In: *Energy and Buildings* 55, pages 668–673.

Sodha, MS, B Kaur, A Kumar and NK Bansal (1986). 'A comparison of the admittance and Fourier methods for predicting heating/cooling loads'. In: *Solar Energy* 36.2, pages 125–127.

Solar Energy Laboratory (2009). *TRNSYS 17 Manual vol. 04 – MathematicalReference*. User Manual. Madison, Wisconsin, USA: University of Wisconsin-Madison.

Stamper, E (1977). 'Weather Data'. In: *ASHRAE Journal* 47.

Steup, M (2014). 'Epistemology'. In: *The Stanford Encyclopedia of Philosophy*. Edited by EN Zalta. Spring 2014.

Struck, C, JLM Hensen and P Kotek (2009). 'On the Application of Uncertainty and Sensitivity Analysis with Abstract Building Performance Simulation Tools'. In: *Journal of Building Physics* 33.1, pages 5–27.

Su, F, YJ Huang, T Xu and C Zhang (2009). 'An evaluation of the effects of various parameter weights on typical meteorological years used for building energy simulation'. en. In: *Building Simulation* 2.1, pages 19–28.

Sulaiman, H and F Olsina (2014). 'Comfort reliability evaluation of building designs by stochastic hygrothermal simulation'. In: *Renewable and Sustainable Energy Reviews* 40, pages 171–184.

Sullivan, L and V Payne (2012). *The RIBA Guide to Sustainability in Practice*. Technical report. London.

Sun, Y, L Gu, CFJ Wu and G Augenbroe (2014). 'Exploring HVAC system sizing under uncertainty'. In: *Energy and Buildings* 81, pages 243–252.

Sun, Y, H Su, CFJ Wu and G Augenbroe (2015). 'Quantification of model form uncertainty in the calculation of solar diffuse irradiation on inclined surfaces for building energy simulation'. In: *Journal of Building Performance Simulation* 8.4, pages 253–265.

Sustainable Energy Research Group (2008). *Climate Change Weather File Generator for the UK*.

– (2013). *Climate Change World Weather File Generator for World-Wide Weather Data – CCWorldWeatherGen*.

Thalfeldt, M, E Pikas, J Kurnitski and H Voll (2013). 'Facade design principles for nearly zero energy buildings in a cold climate'. In: *Energy and Buildings* 67, pages 309–321.

The MathWorks, Inc. (2015). *MATLAB*. Natick, Massachusetts, United States.

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

Tian, W and P de Wilde (2011a). 'Thermal building simulation using the UKCP09 probabilistic climate projections'. In: *Journal of Building Performance Simulation* 4.2, pages 105–124.

– (2011b). 'Uncertainty and sensitivity analysis of building performance using probabilistic climate projections: A UK case study'. In: *Automation in Construction* 20.8, pages 1096–1109.

Tso, GKF and KKW Yau (2007). 'Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks'. In: *Energy* 32.9, pages 1761–1768.

Tuohy, P, S Roaf, F Nicol, M Humphreys and A Boerstra (2010). 'Twenty first century standards for thermal comfort: fostering low carbon building design and operation'. In: *Architectural Science Review* 53.1, pages 78–86.

Üner, M and A İleri (2000). 'Typical weather data of main Turkish cities for energy applications'. en. In: *International Journal of Energy Research* 24.8, pages 727–748.

Van Gelder, L, H Janssen, S Roels, G Verbeeck and L Staepels (2013). 'Effective and robust measures for energy efficient dwellings: probabilistic determination'. In: *Proceedings of BS 2013*. Chambéry, France: IBPSA.

Van Hooff, T, B Blocken, JLM Hensen and HJP Timmermans (2015). 'Reprint of: On the predicted effectiveness of climate adaptation measures for residential buildings'. In: *Building and Environment*. Special Issue: Climate adaptation in cities 83, pages 142–158.

Van Paassen, AH and QX Luo (2002). 'Weather data generator to study climate change on buildings'. en. In: *Building Services Engineering Research and Technology* 23.4, pages 251–258.

Vartiainen, E, K Peippo and P Lund (2000). 'Daylight optimization of multifunctional solar facades'. In: *Solar Energy* 68.3, pages 223–235.

Vignola, F and DK McDaniels (1993). 'Value of Long-term Solar Radiation Data'. In: pages 468–473.

Wan, KK, DHW Li, W Pan and JC Lam (2012). 'Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications'. In: *Applied Energy* 97, pages 274–282.

Ward, RD (1905). 'The Climatic Zones and Their Subdivisions'. In: *Bulletin of the American Geographical Society* 37.7, pages 385–396.

Weisstein, EW (2015). *Regression*. en.

Westphal, FS and R Lamberts (2004). 'The use of simplified weather data to estimate thermal loads of non-residential buildings'. In: *Energy and Buildings*. Performance Simulation for Better Building Design 36.8, pages 847–854.

Wilcox, S and W Marion (2008). *Users' Manual for TMY3 Data Sets*. Technical report.

Wilcox, S (2012). *National Solar Radiation Database 1991–2010 Update: User's Manual*. Technical report NREL/TP-5500-54824. Golden, CO, USA: NREL, National Renewable Energy Laboratory.

Wold, HA (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist & Wiksell.

Wong, NH, KP Lam and H Feriadi (2000). 'The use of performance-based simulation tools for building design and evaluation - a Singapore perspective'. In: *Building and Environment* 35.8, pages 709–736.

Wong, SL, KKW Wan, DHW Li and JC Lam (2012). 'Generation of typical weather years with identified standard skies for Hong Kong'. In: *Building and Environment* 56, pages 321–328.

Wood, M, M Eames and P Challenor (2015). 'A comparison between Gaussian Process emulation and Genetic Algorithms for optimising energy use of buildings'. In: *Proceedings of BS 2015*. Hyderabad, India: IBPSA.

World Climate Research Programme (2015). *CORDEX*.

World Commission on Environment and Development (1987). *Our common future*. Oxford; New York: Oxford University Press.

Wu, S and JQ Sun (2012). 'Multi-stage regression linear parametric models of room temperature in office buildings'. In: *Building and Environment* 56, pages 69–77.

Yan, J, YJ Kim, KU Ahn and CS Park (2013). 'Gaussian process emulator for optimal operation of a high rise office building'. In: *Proceedings of 13th International Building Performance Simulation Association Conference*.

Yang, C, H Li, Y Rezgui, I Petri, B Yuce, B Chen and B Jayan (2014). 'High throughput computing based distributed genetic algorithm for building energy consumption optimization'. In: *Energy and Buildings* 76, pages 92–101.

Yang, H and L Lu (2004). 'Study of typical meteorological years and their effect on building energy and renewable energy simulations'. English. In: volume 110 PART II, pages 424–431.

Yang, L, JC Lam and J Liu (2007). 'Analysis of typical meteorological years in different climates of China'. In: *Energy Conversion and Management* 48.2, pages 654–668.

Yang, L, KKW Wan, DHW Li and JC Lam (2011). 'A new method to develop typical weather years in different climates for building energy use studies'. In: *Energy* 36.10, pages 6121–6129.

Zhao, Hx and F Magoulès (2012). 'A review on the prediction of building energy consumption'. In: *Renewable and Sustainable Energy Reviews* 16.6, pages 3586–3592.

Zhao, J, R Plagge, NMM Ramos, ML Simões and J Grunewald (2015). 'Concept for development of stochastic databases for building performance simulation – A material database pilot project'. In: *Building and Environment* 84, pages 189–203.

Zhu, M, Y Pan, Z Huang and P Xu (2016). 'An alternative method to predict future weather data for building energy demand simulation under global climate change'. In: *Energy and Buildings* 113, pages 74–86.

Glossary

Symbols

n_{boot} Number of bootstrap samples. 265

n_{sim} Number of simulations of the Seasonal Auto-Regressive Moving Average (SARMA) model. 265

A

ACF Auto-Correlation Function 108, 114, 115, 265

AIA American Institute of Architects 265

AIC Akaike Information Criteria 97, 107, 109, 265

ANOVA Analysis of Variance 159, 160, 265, 272

AR Auto-Regressive 44, 78, 83–86, 107–110, 213, 214, 265

ARCH Auto-Regressive Conditional Heteroscedasticity 111, 215, 265

ARD automatic relevance determination 170, 172, 225–227, 238, 265

ARIMA Auto-Regressive Integrated Moving Average 107, 214, 265

ARMA Auto-Regressive Moving Average 85, 107–109, 111, 135, 136, 203, 212, 214, 265

ASCE American Society of Civil Engineers 265

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers 17, 35, 36, 43, 45, 72, 75, 78, 90, 123, 124, 127, 129, 139, 216, 265

atmospheric pressure Atmospheric pressure at a given station is usually stated in Pascals (Pa) or hPa (hecto-Pascals) in this document. The atmospheric pressure is less important for building energy simulation and more for reducing the loss of information in converting from Relative Humidity (RH) to Humidity Ratio (W) and vice-versa. If the actual value of pressure at a given location and time was unavailable, the U.S. Standard Atmosphere was used in the calculations, which is not ideal. The equations for these conversions come from *ASHRAE Handbook: Fundamentals*. 78, 265

ATMPR Atmospheric Pressure 265, *see* atmospheric pressure

B

BEE Bureau of Energy Efficiency [Govt. of India] 265

BIC Bayesian Information Criteria 97, 107, 109, 265

BIM Building Information Modelling 68, 265

BLUP Best Linear Unbiased Predictor 252, 265

Bootstrap The term bootstrap (sometimes used as a verb *bootstrapping*) refers to a range of techniques "used to estimate the standard error, the bias and the confidence interval of a parameter (or more than one parameter)" (Dodge 2008). More details can be found in Davison and Hinkley (1997), NIST (2013) and Politis (1998). 265

BPS Building Performance Simulation 6–8, 19, 49, 66, 67, 69, 96, 265, *see* building performance simulation

BREEAM® Building Research Establishment Environmental Assessment Methodology 48, 265

building envelope Used interchangeably with 'facade', it refers to those elements of a building that form an interface between a building and its environment. This usually is taken to include the walls, roof, and floor. In our analyses, we often ignore the floor, not including it, for example, in the factorial experiment of U-values. 265, 271, *see* facade

building performance simulation Building Performance Simulation (BPS), or building simulation for short, is a phrase used to describe a set of tools and methods to model the flow of heat and mass in and around buildings. The term BPS is generally taken to encompass thermal, daylight, fluid (air and pollutants), and acoustics simulation. In the context of this document, BPS will refer to thermal simulation unless otherwise specified. 25, 49, 93, 265

C

CAD Computer-Aided Design 68, 265

CDD Cooling Degree Day 25, 265

CDF Cumulative Density Function 73, 74, 77, 79, 86, 265, 279

CI Confidence Interval 265, *see* confidence interval & confidence level

CIBSE The Chartered Institution of Building Services Engineers 72, 75, 78, 90, 265

confidence interval "A confidence interval [abbreviated as CI] is any interval constructed around an estimator that has some probability of containing the true value of the corresponding parameter of a population" (Davison 2003). The extents of a confidence interval are governed by the confidence level. For example, a confidence level of 0.05 corresponds to a CI of 95%. If, from a given sample, we calculate that the upper and lower limits of a CI for the sample mean $\bar{x}$ are k_1 and k_2 respectively, then it **does not** follow that the there is a 95% chance of finding the mean of the *population* in the given interval. Once a sample is taken, the true parameter may or may not fall in the given interval. Rather, if the experiment were repeated several times, and a CI calculated from each sample ($\{k_3, k_4\}, \ldots \{k_{N-1}, k_N\}$), then 95% of the confidence intervals so obtained will contain the true value of μ. Or, on average, 19/20 confidence intervals with coverage 0.95 or 95% will contain the true value of the parameter in question. In practice, confidence intervals are approximate, since they are calculated from the data at hand. They can be calculated exactly if the underlying distribution is known or assumed (e.g., Normal). (Davison 2003; Dodge 2008) 7, 8, 23–25, 65, 86, 194, 195, 198, 223, 224, 265, 269, 280, *see* confidence level, confidence interval, prediction interval & credible interval

confidence level Analogous to a confidence interval, a confidence level is the probability of finding the true value of a population parameter in the corresponding confidence intervals constructed around an estimator from several samples (Dodge 2008). Commonly used confidence levels include $\frac{\alpha}{2} = 0.025$, $\frac{\alpha}{2} = 0.05$, corresponding to confidence intervals of $(1-\alpha) \times 100\% = 95\%$ and 90% respectively. 223, 265, *see* confidence interval

correlation coefficient The correlation coefficient is a measure of the strength of the relation between two random variables. The dependence can be linear, in which case it may be evaluated using, e.g. Pearson's correlation coefficient (Pearson 1901). Pearson's correlation coefficient is the ratio between the covariance of two vectors and the product of their individual variances. 149, 265

credible interval The Bayesian analogue to a confidence or prediction interval, it demarcates a belief about the true value of a quantity. Credible intervals are also stated in terms of confidence levels. 265, *see* confidence interval & confidence level

CSI Climate Severity Index 36, 37, 265

CSV Comma-Separated Values (file) 265

CWEC Canadian Weather for Energy Calculations 75, 265

D

Degree Day A Degree Day is a calculated quantity that aggregates the variation of temperature from a given 'balance point' with the amount of time for which the variation occurred. For example, assume that a given building does not need heating or cooling when the outside temperature is 15^{o}C, i.e. its balance point. If the outside temperature is 20^{o}C for 6 hours continuously, then the number of *cooling* degree days accumulated is $\Delta T * t_{days} = (20-15) * 0.25 = 1.25$. Then if the temperature drops to 10^{o}C for a further 12 hours, the number of *heating* degree days accumulated is $\Delta T * t_{days} = (15-10) * 0.5 = 2.50$. It is important to note that, since the effect of thermal storage is ignored, heating and cooling degree days do not cancel each other out. That is, one would need cooling above the balance point and heating below it for distinct time steps. 35, 36, 38, 39, 42, 44, 265

DFT Discrete Fourier Transform 210, 265

DHI Diffuse Horizontal Irradiation 96, 121, 265

DNI Direct Normal Irradiation 74, 79, 96, 121, 265

DoE Design of Experiments 265

DRY Design Reference Year 27, 41, 42, 72, 73, 265, 279

DSA Differential Sensitivity Analysis 64, 265

DSY Design Summer Year 40, 41, 76, 80, 81, 265

DWC Design Weather Conditions 90, 265

E

eCDF empirical Cumulative Density Function 74, 123, 125–127, 129, 130, 133, 134, 217, 219, 220, 265

Energy Plus A whole building energy simulation program that engineers, architects, and researchers use to model energy and water use in buildings (NREL and USDOE 2015). 17, 91, 160, 209, 254, 265

ESP-r A modelling tool for building performance simulation developed by the Energy Systems Research Unit at University of Strathclyde, Glasgow (ESRU 2015). 61, 265

ESRU Energy Systems Research Unit 265

EUI Energy Use Intensity 10, 89, 90, 145, 265

F

facade Spelt as both facade and façade, this term is used interchangeably with the word building envelope in this work. 265, *see* building envelope

FFT Fast Fourier Transform 265

Form Factor The ratio between the volume of the building and the area of the (vertical) envelope. 265

FS statistic Finkelstein-Schafer statistic 42, 44, 73, 74, 77, 265

Full Factorial Experiment A full, or complete, factorial experiment is one in which all the possible levels of a given parameter (factor) are investigated in such a way as to reveal the main effects and interactions. This is done by combining the different levels of each factor in every possible way. For our purposes, we use an additive linear model/relationship between the independent and dependent variables. Dodge (2008) 265

G

G-value Roughly, the amount of solar radiation that enters through a window glass (transmittance). G-value is used primarily in Europe, and is analogous to the Solar Heat Gain Coefficient (SHGC) from the US. SHGC commonly includes the window frame, sash, etc. 265, *see* SHGC

GA Genetic Algorithm 70, 265

GARCH Generalised Auto-Regressive Conditional Heteroscedasticity 111, 215, 265

Gaussian Process regression Gaussian Process regression is a supervised kernel-based machine-learning method. Ebden (2008) states that it is a method where the "data has to do the talking", though it is not completely free-form, assuming as it does that the data is generated "...throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution." Rasmussen and Williams (2006) define a Gaussian Process as "a collection of random variables, any finite number of which have a joint Gaussian distribution". 9, 23–25, 37, 67, 70, 144, 145, 155, 158, 167–169, 171–173, 175, 188, 190–192, 194, 195, 197, 203, 225, 227, 229, 238, 265

GCM Global Climate Model 39, 101, 265

GEV Generalised Extreme Value [distribution] 265

GHG Green House Gas 3, 65, 101, 265, 276

GHI Global Horizontal Irradiation 74, 76, 77, 79, 80, 85, 96, 120–122, 128, 194, 249, 265

GLM Generalised Linear Model 144, 159, 162, 265

GLMM Generalised Linear Mixed-Effects Model 144, 156, 158, 159, 162, 265

GP Gaussian Process 24, 70, 144, 155, 156, 168, 169, 171, 225, 228, 265, 272, 278, *see* Gaussian Process regression

GRIHA® Green Rating for Integrated Habitat Assessment 48, 265

GUI Graphical User Interfaces 68, 84, 265

H

HDD Heating Degree Day 40, 265

HVAC Heating, Ventilation, and Air Conditioning 3, 39, 45, 46, 52, 62, 71–73, 86, 89, 145, 265

I

IBPSA International Building Performance Simulation Association 265

IEA International Energy Agency 265

IGBC Indian Green Building Council 265

Internal Heat Gain Defined in this work as the sum of (usually sensible) heat gains from people, equipment, and lights. Technically, internal heat gains should include latent gains, but we do not consider them in the context of this work. 29, 44, 151, 204, 265

IPCC Intergovernmental Panel on Climate Change 101, 265, 276

IQR Inter-Quartile Range 247, 265

ISHRAE Indian Society of Heating, Refrigerating and Air-Conditioning Engineers 265

IWEC International Weather for Energy Calculations 75, 155, 265

K

Kriging Kriging is a technique borrowed from geostatistics, originally developed by G. Matheron based on the work of D. G. Krige (Kleijnen 2009). In this thesis, we use the term GP instead of Kriging since that is the exact technique we are working with. 91, 169, 265

Kruskal-Wallis test This is a non-parametric test which tests if all k samples in a given set of samples come from the same population (Dodge 2008). The null hypothesis is that the k treatment medians are identical, against the alternative that at least one of them is different. This test is a non-parametric version of the classical one-way Analysis of Variance (ANOVA), since it uses ranks rather than the numeric data for comparison. 265

KS statistic Kolmogorov-Smirnov statistic 73, 265

L

LBNL Lawrence Berkeley National Laboratory 265

leakage In the context of the discrete Fourier transforms used in this thesis, leakage is "the appearance of a non-zero value in the transform at a frequency f because of the presence of a sinusoid at a different frequency f_0 ..." (Bloomfield 2000) 103, 104, 265

LEED® Leadership in Energy and Environmental Design 45, 48, 265

LHS Latin Hypercube Sampling 60, 62, 265

LMM Linear Mixed-Effects Model 159, 161, 165, 188, 189, 265

M

MA Moving Average 78, 86, 108, 109, 214, 265

MAD Median Absolute Deviation 265

Mahalanobis distance The Mahalanobis distance is a multidimensional generalisation of how many standard deviations away X_i is from the mean of a (empirical) distribution D. That is, the distance of a given data point X_i from the expected value of the distribution $E(D) = \mu$, in units of variance σ_D^2. The distance is calculated, in squared units, as

$$d_M^2(X,\mu) = (X-\mu)^t \cdot \Sigma^{-1} \cdot (X-\mu) \quad , \tag{B.15}$$

where $d_M^2(X,\mu)$ is the distance of the given data point from a data set, $(\cdot)^t$ denotes transposition, and Σ is the $p \times p$ covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \mathrm{cov}(X_1,X_2) & \cdots & \mathrm{cov}(X_1,X_p) \\ & \sigma_2^2 & \cdots & \mathrm{cov}(X_2,X_p) \\ \vdots & & \ddots & \vdots \\ \mathrm{cov}(X_p,X_1) & & & \sigma_p^2 \end{bmatrix} .$$

The Mahalanobis distance is unitless, scale-invariant, and takes into account the correlations of a data set. (Dodge 2008; Orlov 2011) 120, 265

MC Monte Carlo 6, 59, 63, 139, 143, 265

MCA Monte Carlo Analysis 8, 64, 195, 265

Mesoscale Meso-scale is a commonly used scale in meteorology implying phenomenon whose range is more than 2 km but less than 2000 km. Building sites are anywhere between a few dozen to hundreds of kilometres from an actual measurement station, so climatic variations can already arise without even considering the effects of the urban fabric. Spatial variance in weather data between stations is a bigger problem in countries where weather data availability is sparse. 91, 265

MLE Maximum Likelihood Estimation 116, 155, 227, 265

MLE Maximum Likelihood Estimate 228, 265

MN METEONORM 71, 82, 83, 91, 155, 265

Monte Carlo simulation Monte Carlo simulation is a well-established method of accounting for the effects of an input whose exact future values are unknown, i.e. a (pseudo-) random input. Dodge (2008) defines Monte Carlo simulation as a "numerical technique for solving mathematical problems... that do not have analytical solutions... [using] random or pseudo-random numbers". This is precisely what we accomplish by creating the synthetic weather files described in this thesis. 11, 59, 265

MRE Median Relative Error 188–190, 265

MSPE Mean Squared Prediction Error 265

MTM Markov Transition Matrices 83, 265

Multivariate Gaussian Distribution Another term for the Multivariate Normal Distribution. 265, *see* Multivariate Normal Distribution

Multivariate Normal Distribution A multivariate Normal distribution, as the name implies, is a generalisation of the Normal distribution to many dimension, say $n \in \mathbb{R}^d | n \leq d$. Santner, Williams et al. (2003) define a multivariate Normal or Gaussian Distribution in terms of an "affine combination of independent and identically distributed standard normal random variables". An *affine combination* is simply a linear combination in which the coefficients sum to 1. Suppose $\boldsymbol{Z} = \{Z_1, Z_2, \ldots, Z_r\}$ is a set of independent, identically distributed random variables, such that $Z_i \sim \mathcal{N}(0,1)$. Let $\boldsymbol{L}$ be some $m \times r$ real matrix and μ an $m \times 1$ real vector. Then the $m \times 1$ vector of vectors

$$\boldsymbol{W} = \boldsymbol{LZ} + \boldsymbol{\mu} = \{W_1, W_2, \ldots, W_m\} \tag{B.16}$$

is said to have a *multivariate normal distribution.* The first and second moments of $\boldsymbol{W}$ are

$$\boldsymbol{\mu} = E\boldsymbol{W}, \mathrm{Cov}(\boldsymbol{W}) = E\big((\boldsymbol{W} - \boldsymbol{\mu})(\boldsymbol{W} - \boldsymbol{\mu})^T\big) = \boldsymbol{LL}^T. \tag{B.17}$$

265, 274, *see* Multivariate Gaussian Distribution

N

NCDC National Climatic Data Center 72, 209, 221, 265

NLM Normal Linear Model 145, 154, 157–159, 161, 163–166, 188, 189, 265

NREL National Renewable Energy Laboratory 74, 265

NSRDB [United States] National Solar Radiation Database 221, 265

NZEB Net Zero-Energy Building 53, 265

P

PACF Partial Auto-Correlation Function 108, 115, 265

PC Principal Component 44, 76, 149, 252, 253, 265, *see* Principal Component Analysis

PCA Principal Component Analysis 44, 54, 148, 149, 225, 252, 253, 265, *see* Principal Component Analysis

PDF Probability Density Function 58, 59, 111, 123, 125, 126, 253, 265

Phase-Change Material All materials undergo phase transitions at some combination of temperature and pressure. In the context of building physics though, phase-changing materials refers to specific materials which have a high heat of fusion, transition from liquid to solid (and vice-versa) at some specific temperature near room temperature, and preferably do not undergo great changes of volume. They are usually used to give a building Thermal Mass, and include water, paraffins, inorganic salt hydrates. 62, 255, 265, 279

PI Prediction Interval 265

PMV Predicted Mean Vote 46, 265

PPD Percentage People Dissatisfied 46, 265

prediction interval An interval analogous to the variability interval, but this time on a predicted observation (i.e., at some new query point). Prediction intervals are also stated in terms of confidence levels. 161, 167, 172–174, 178, 179, 182, 183, 223–225, 234–237, 243, 244, 265, *see* confidence interval & confidence level

Principal Component Analysis Principal Component Analysis is a mathematical transformation of a multidimensional data space which allows the reduction of variables needed to describe a space. The original variables are combined linearly into new variables called Principal Components. "The first principal component is required to have the largest possible variance... The second component is computed under the constraint of being orthogonal to the first component and to have the next largest possible variance" (Abdi and Williams 2010), and so on. So long as the data set is joint normally distributed, each of these components is orthogonal to the others. The outcome of this transformation is a new set of variables which are linear combinations of the old ones. Each of these new variables has a corresponding eigenvalue and ‘cumulative explained variance’, which are used to pick the principal components that should be retained. Not all of the principal components need to be kept so a low-dimensional picture of the original dataset can be obtained. The concept was first proposed by Pearson (1901). In this work, we use the MATLAB implementation `pca` (The MathWorks, Inc. 2015). 76, 149, 265

PSD Power Spectral Density 103, 106, 210, 212, 265

PV PhotoVoltaic 8, 265

R

R-value The R-value of a building component, or thermal resistance, is the reciprocal of U-value. Usually expressed in $m^2 - K/W$. 64, 265, 280, *see* U-value

RBF radial basis function 225, 265

RCM Regional Climate Model 39, 40, 101, 265

RCP Representative Concentration Pathway 97, 126, 132, 265, 276, *see* Representative Concentration Pathway

RE Relative Error 265

Regression Analysis Regression analysis is a method used to obtain a mathematical relationship between two or more quantities. The aim is to be able to estimate one (dependent) variable as a function of another (independent) variable, or several others (Dodge 2008). Regression involves fitting a curve“through a set of points using some goodness-of-fit criteria" (Weisstein 2015). 69, 265

Representative Concentration Pathway According to the synthesis of the Intergovernmental Panel on Climate Change (IPCC)’s Fifth Assessment Report (IPCC 2014b), “...anthropogenic GHG emissions are mainly driven by population size, economic activity, lifestyle, energy use, land use patterns, technology and climate policy. The Representative Concentration Pathways (RCPs), which are used for making projections based on these factors, describe four different 21st century pathways of Green House Gas (GHG) emissions and atmospheric concentrations, air pollutant emissions and land use. The RCPs include a stringent mitigation scenario (RCP2.6), two intermediate scenarios (RCP4.5 and RCP6.0) and one scenario with very high GHG emissions (RCP8.5). Scenarios without additional efforts to constrain emissions (‘baseline scenarios’) lead to pathways ranging between RCP 6.0 and RCP 8.5 (Figure SPM.5a [in the text]). RCP2.6 is representative of a scenario that aims to keep global warming likely below 2°C above pre-industrial temperatures. The RCPs are consistent with the wide range of scenarios in the literature as assessed by WG III [Working Group 3]." 15, 96, 265

RH Relative Humidity 76, 78, 96, 100, 103, 104, 106, 111, 112, 115, 125, 126, 128, 140, 141, 194, 216, 217, 219, 251, 265

RMSE Root Mean Square Error 155, 170, 171, 188–190, 265

Roof Ratio The ratio between the area of the roof and the area of the (vertical) envelope. 265

S

SA Sensitivity Analysis 60, 62, 63, 65, 67, 69, 144, 265

SAR Seasonal Auto-Regressive 135, 265

SARMA Seasonal Auto-Regressive Moving Average 97, 108, 109, 111–113, 116, 118, 124, 136, 139, 265, 267

SD Standard Deviation 265

SE Standard Error 265

Self-shading For our purposes, self-shading is defined as the shading of the transparent elements of a building's envelope (glazing) by elements of the envelope itself. So, for example, shading from neighbouring buildings is not counted as 'self'-shading. 265

SHGC Solar Heat Gain Coefficient 44, 265, 271, *see* G-value

ShRY Short Reference Year 78, 265

SIA Swiss Society of Engineers and Architects 6, 265

SMA Seasonal Moving Average 110, 135, 265

SQ Sensitivity Quantification 66, 265, *see* SA

SqE squared exponential function 171, 172, 225–227, 229, 265

SRC Standardised Regression Coefficients 62, 265

SSA Stochastic Sensitivity Analysis 64, 265

standard deviation A measure of dispersion, it is the positive square root of the variance. The standard deviation of a parameter estimated from several sample is known as its standard error (Dodge 2008). Unless otherwise specified, standard deviation implies the dispersion of a *raw value* in this thesis. We almost always talk about *sample* standard deviations, and the question of knowing the population standard deviation is somewhat theoretical. The sample standard deviation is

$$S = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1}}, \tag{B.18}$$

where n is the number of observations, $\bar{x}$ is the sample mean, and i is the index of observations. The denominator could also be n, used when estimating the standard deviation from large samples. By default, the standard deviation calculations in this thesis will use $n-1$. 73, 157, 167, 169, 170, 173, 265, 280, *see* standard error & variance

standard error The standard error is the estimated standard deviation of some parameter estimated from a sample. From each sample, we can calculate some statistic like the mean. If more samples are taken from the same population, then the mean calculated from each sample will be different. An estimate of the standard deviation of the population of these means would be the standard error of the mean. 265, *see* standard deviation

stationarity The simplest definition of a stationary process, say a time series g_t or a process $g(t)$, is one that has a constant mean $E(g_t) = \mu$, and a stable variance $\mathrm{Var}(g_t) = \sigma$. A stable variance implies that $Cov(g_t, g_{t+k}) = \sigma(k)$, i.e., the covariance (function) $\sigma(k)$ depends only on time lag k and not on absolute time t. This is a weak, or second-order, definition of stationarity, but sufficient for most applications. An interesting aside is that if $g(t)$ is generated by an underlying GP, then fulfilment of the weak stationarity conditions implies strict stationarity (Christensen 1991). We will use these processes in chapter 4, assuming that the energy consumption of a building is well modelled by a GP. 98, 210, 265

surface energy balance The surface energy balance consists of three terms and a residual: $Q - H - E - B = 0$. Q is the net radiation, dependent on latitude and cloudiness; E is the turbulent flux of latent heat, i.e. the energy used for changing the phase of moisture present at the location; H is the turbulent flux of sensible heat, i.e. the energy used to increase the temperature of the surrounding air; and, B is the flux that is absorbed by the ground in addition to the errors in the other three terms, i.e. the residual. 33, 265

SuRY Summer Reference Year 41, 265

SVM Support Vector Machine algorithms are a type of supervised kernel-based machine-learning algorithms, similar to GP. They are also used for machine learning tasks such as classification and regression. 67, 77, 265

SVM Support Vector Machine 67, 265, *see* SVM

synoptic “Pertaining to or forming a synopsis; furnishing a general view of some subject; spec. depicting or dealing with weather conditions over a large area at the same point in time.” (Oxford English Dictionary 2016) 79, 265

T

TDB Dry Bulb Temperature 41, 42, 74, 76–78, 80, 84, 85, 95, 98, 100, 102–105, 110–112, 114, 115, 118, 120–122, 125, 126, 128, 132, 140, 149, 150, 159, 194, 197, 211, 217, 219, 249, 250, 265

TDP Dew Point Temperature 74, 76, 80, 84, 85, 96, 149, 250, 265

Thermal Mass The thermal mass, or thermal density, of a material refers to its ability to store heat. In the context of buildings, this provides a thermal “inertia” that dampens the fluctuations of temperature that would be otherwise induced by changing weather conditions. Most materials have some thermal mass, though the ones that are ordinarily considered in building simulation include concrete, wood (especially furniture), earth, and Phase-Change Materials. See Schröder and Gawron (1981) for a discussion of latent heat storage, including Phase-Change Materials. 53, 151, 253, 265, 275

TMY Typical Meteorological Year 27, 42, 71, 73–81, 89, 92, 96, 97, 113, 120–127, 131, 132, 136–138, 141, 155, 265, 279

TMY2 Typical Meteorological Year - Version 2 265

TMY3 Typical Meteorological Year - Version 3 71, 92, 265

TPCY Typical Principal Component Year 76, 77, 265

TRY Test Reference Year 40, 41, 72, 74, 76, 78, 80, 265, 279

TXT Text (file) 265

typical year Typical year files exist in various avatars, known variously as Test Reference Year (TRY), Typical Meteorological Year (TMY), standard years, example years, Design Reference Year (DRY), etc. A typical year is a whole year of hourly (or finer) weather data compiled for simulating building performance or renewable energy production. Typical year datasets are not designed to contain extremes, but instead to capture seasonal variability and the characteristics of the local climate (3TIER 2011). Common algorithms to select typical years include the TMY algorithm from Marion and Urban (1995) and the TRY algorithm from Levermore and Parkinson (2006). The former selects months based on the closeness of their Cumulative Density Function (CDF) to the CDF of the overall data. The latter selects months based on the closeness of their mean to the mean for that month throughout the period of record. 23, 40, 42, 72, 74, 80, 81, 93, 96, 127, 135, 137, 141, 145, 150, 156, 158, 209, 265

U

U-value The U-value of a building component, or overall thermal transmittance, is a simplified material thermal property that is considered to be fixed and isotropic in building simulation. It is composed of the conductivity and thickness of each solid material, and the combined radiative and convective thermal resistance of the innermost surface, outermost surface, and any cavities. U-value is the reciprocal of R-value or resistance. Both U- and R-Values are used in lumped-heat-capacity models, where the internal temperature gradient, brought about by the 'internal' resistance, is neglected (Holman 1972). Usually expressed in $W/m^2 - K$. 28, 44, 45, 51, 53, 58, 59, 64, 154, 157, 197, 265, 276, *see* R-value

U.S. Standard Atmosphere Owen (2013) defines the U.S. standard atmosphere, at sea level, as having a temperature of 15°C and barometric pressure of 101.325 kPa. 265, 267

UA Uncertainty Analysis 55, 56, 60, 63, 67, 69, 265

UHI Urban Heat Island 39, 137, 265

UQ Uncertainty Quantification 60, 63, 66, 265, *see* UA

USDOE United States Department of Energy 146, 153, 161, 163, 164, 172, 188–190, 254, 265

USGBC United States Green Building Council 265

V

variability interval An interval over which the raw values of a sample vary. We distinguish between confidence intervals, which are constructed for a statistical quantity, and variability intervals, which indicate the range of values of a raw quantity. 55, 143, 194, 195, 198, 223, 265

variance Variance is the square of standard deviation. 86, 99, 265, *see* standard deviation

W

W Humidity Ratio 96, 265

Window-to-Floor Ratio The ratio between the area of windows (transparent elements) and the area of the floor or footprint of the building. 149, 265

Window-to-Wall Ratio The ratio between the area of windows (transparent elements) and the area of the wall. 149, 197, 253, 265

WMO World Meteorological Organization 9, 265

X

XMY eXtreme Meteorological Year 76, 265

PARAG RASTOGI

Av de la Chablière 35b
CH-1004 Lausanne (Switzerland)
☎ +41 78 943 00 27
✉ rastogi.parag@gmail.com
Born: 27.08.1988, Nationality: Indian

I am passionate about buildings, infrastructure, and climate. When not working on challenging trans-disciplinary problems in building physics, I am to be found reading about or teaching energy, development, the environment, and policy.

EDUCATION

2012 – PRESENT | **PhD in Civil Engineering, École Polytechnique Fédérale de Lausanne**
Interdisciplinary Laboratory of Performance-Integrated Design.
Advisor: Professor Marilyne Andersen
Thesis – On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption

2010 – 2011 | **Master of Science in Civil Engineering**, **Purdue University**
Charles C. Chappelle Fellowship: awarded for the furtherance of post-graduate research, based on character and intellectual ability.

2006 – 2010 | **Bachelor of Science in Civil Engineering, Purdue University**
Graduated with University and Departmental Honours: programmes for students with high grades who wish to supplement their regular coursework with challenging projects and research.
Other awards:
- Clark Scholarship and Miles Scholarship, *School of Civil Engineering*
- Purdue Academic Success Award, *Department of Admissions*
- Edward J. Cox Memorial Scholarship, *Institute of Transportation Engineers, Indiana, USA*
- GANA Undergraduate Scholarship, *Glass Assoc. of North America*

University College London (2008-09) – Exchange student

WORK AND TEACHING EXPERIENCE

2012 – PRESENT | **Doctoral Assistant**, École Polytechnique Fédérale de Lausanne
Teaching and assistance for several Bachelors and Masters courses
- Technical ecology of human communities
- Renewable Energy and Solar Architecture in Davos
- Performance Studies: Towards Solar Decathlon
- Comfort and Architecture: Sustainable Strategies
- Solar Decathlon Workshop

2014 | Supervised MSc thesis (G. Chinazzo, see publications)

2012-2014 | Supervised Masters level semester projects focussing on renewable energy and building simulation

2010 – 2011	**Research Fellow**, Purdue University School of Civil Engineering Developed code to simulate experimental testing rooms for daylighting and energy.
MAY-AUG 2011	**Trainee Engineer**, Lutron Electronics Co., Inc. Developed a novel daylight-based control system for venetian blinds. Validated the system with experiments and simulation.
MAY-SEPT 2008	**Engineering Intern**, Parsons Brinckerhoff International Managed the resolution of design and contractual issues with the design-build contractor. Reviewed civil and general drawings. Managed an Access® database of contractual communication.

LANGUAGES

ENGLISH	Excellent	*Native*
HINDI/ URDU	Excellent	*Native*
FRENCH	Good	*B1-B2, 4 years in Switzerland*
GERMAN	Fair	*A1*

SOFTWARE AND PROGRAMMING

DATA ANALYSIS AND STATISTICS	MATLAB (expert), Python (basic), R (basic), MS Excel (expert), MS Access (basic)
BUILDING SIMULATION	EnergyPlus (expert), DesignBuilder (expert).

EXTRA-CURRICULAR ACTIVITIES

2014 – 2015	**PhD Representative**, Doctoral Program in Civil and Environmental Engineering Organised professional and social events for the students.
2014 – PRESENT	**LEDSafari®** (a Switzerland-based social start-up, www.ledsafari.com) Created and edited teaching content about renewable energy and physics. Led workshops in India and Switzerland.
2009 – 2011 2009 – 2010	**Ambassador,** Purdue University School of Civil Engineering and Purdue University Study Abroad Programs Office Represented the offices at fairs and events. Recruited new students.
2009 – 2010	**Purdue University Freshman Honours Engineering Program** Mentored students based on my own experience in the program.

CERTIFICATIONS

2010 – PRESENT	**Engineering Intern** National Council of Examiners for Engineering and Surveying (NCEES), USA
2011 – 2015	**LEED® Green Associate** US Green Building Council, Washington D.C., USA

HOBBIES

I am a voracious reader of fiction, current affairs, poetry, and philosophy. I am often in the mountains to hike or ski with friends, or on the road for a run or bike ride. My other pastimes include travel, films, cooking, and writing poetry.