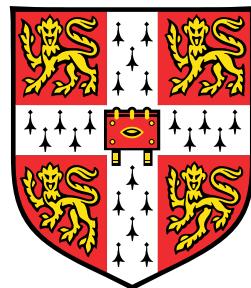


The importance of data collection to support decision making in energy systems



Max Langtry

Supervised by
Prof. Ruchi Choudhary

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

To my family.

I love you.

*This may not seem
Very important, I know.
But it **is**. So I'm bothering
Telling you so.*

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Max Langtry
September 2025

Use of Generative AI

This thesis and much of the code for the experiments was written in VScode with Copilot as an inline AI assisstant, using GPT-4o. It was also proofread by Gemini 2.5 Pro, and the Abstract and Introduction were narrated by ElevenLabs.

These models have been used responsibly. All the ideas in this thesis are my own, but the work has been AI accelerated.



Acknowledgements

There's only one name on the front of a thesis. But it takes a village to write one.

I've had the privilege of getting to know so many amazing people during my time at Cambridge. I really believe that they're the best thing about this university. Though the swoosh of a gown as you stride across a college courtyard does run a close second. Without their support, encouragement, and advice, I wouldn't have even attempted a PhD, let alone finished one.

And so I will attempt, and fail, to communicate the depth of my gratitude to all those who have helped me on this journey.

First and foremost, I owe so much of my progress over this PhD, both in research and beyond, to the patience and guidance of my supervisor, Ruchi. Thank you for so expertly finding that delicate balance of supporting me to do my best, and encouraging me to do even better. I couldn't have hoped for a better supervisor.

One of Ruchi's many talents is that she gathers wonderful people around her. I've learnt much from the EECi team, and had lots of fun doing it, including a truly delightful month in Stanford. I've also had the chance to work with other fantastic researchers outside of EECi: Zoltan Nagy, Kingsley Nweye, Wonjun Choi, Jie Lu, and Fabian Raisch. Collaborating with them has been rewarding, interesting, and has prompted me to consider if the Cambridge way of doing things is in fact the only way.

I wouldn't have even considered doing a PhD without the support and encouragement of John Miles. Thank you for being so generous with your time and advice, and giving me the opportunity to discover for myself the joy of research. I'm also very grateful to Domenic Di Francesco for patiently guiding me through my initial stumblings with the Value of Information, and persevering with me as I figured out how to say what I actually mean. My journey into energy research began with David MacKay's quintessential book, Sustainable Energy – Without the Hot Air. It continues to inspire me to look at energy more clearly, and communicate more succinctly.

Through my time at Emmanuel College, Alexandre Kabla has been an incredible mentor and supporter, always finding the time to help me no matter how busy term got.

And I would like to give a special thanks to the ever amazing Emmanuel Engineers. They have been an absolute joy to teach, always managing to teach me something in the process, and keeping me grounded.

To my friends. You know who you are you lovely bunch. Thank you for the coffees, the gigs, the runs, the laughs, the scran, the wittical silliness, the thoughtful evenings, and the time shared. I couldn't imagine my time in Cambridge without you. There would've been fewer biscuits.

And to Greg. For the tea.

Lastly, and most importantly, to my family. It's hard to do any justice to how much your love and support means to me and has helped me. Sharing my time at Cambridge with you has made it endlessly more special. Thank you. I love you.

A handwritten signature in black ink that reads "Max".

This research was supported by the Engineering and Physical Sciences Research Council, through the CDT in Future Infrastructure and Built Environment: Resilience in a Changing World, Grant [EP/S02302X/1].

The experiments in this thesis were made possible by the superb service provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council, Grant [EP/T022159/1], and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

Howay pet, tek a pew like, let w' telt ya sumat



Abstract

Energy systems need to be designed, operated, and managed effectively to make decarbonisation affordable. But there are many aspects of energy systems, such as their cost and the operating conditions they will face, that are not known precisely at the time these decisions need to be made. To manage these uncertainties, decision makers have to hedge their choices against the possible outcomes, missing out on the best possible choice for the actual system state. Collecting data can reduce uncertainty, and enable better decisions. But data is itself costly. This raises two crucial questions, “How much does collecting data to reduce uncertainty improve decision making? And are the benefits the data provides to decision making worth its cost?”.

Understanding of how data collection and uncertainty reduction affect decision making in energy systems is needed to determine where data should be used to improve decisions, and where wasting time and resources on low insight data can be avoided. However, no existing studies have numerically determined the data requirements for supporting the design and operation of energy systems, or quantified the benefit of practical data collection for improving decision making in energy systems.

This thesis investigates the impact that data and uncertainty reduction have on the design and operation of energy systems across scales. It aims to demonstrate the importance of considering how data affects decision making, and show that **Value of Information** analysis (**Vol**) provides a clear and rigorous methodology for studying this. It begins by explaining the **Vol** framework, and then extends it to allow the study of complex decision making problems where decision policies are required, and the benefit of retaining optionality as uncertainty reduces. How **Vol** can be used to study varying energy system contexts, and the insights it can provide into the role of data collection for supporting decision making, are then demonstrated via three example decision problems.

Next, the effect of training data on the accuracy of machine learning based forecasting models, and the resulting performance of **Model Predictive Control** in a multi-building energy system is analysed. A simple neural model is found to provide equivalent forecast accuracy to state-of-the-art models, but with better data efficiency and generalisability. Further, using more than 2 years of load data for training provides no significant improvement in forecast

🔊 You can listen to this abstract [here](#).

accuracy, and the accuracy and data efficiency of the models can be simultaneously improved by using change-point analysis to remove non-representative data from the training set. Reusing models achieves prediction accuracies within 10% of the baseline without using any data from the target building, suggesting good forecasts can be made without load data from a building.

The value of using load monitoring data to support the design of a grid constrained district energy system is then quantified. Uncertainty in building load is found to significantly impact both system operating costs ($\pm 30\%$) and the optimal system design ($\pm 20\%$). However, using building monitoring data to improve the sizing of solar-battery systems in the district reduces overall costs by less than 1.5% on average. This saving is less than the cost of obtaining the load data, and so using monitoring is shown to be not economically worthwhile. This provides the first numerical evidence to support the sufficiency of using standard building load profiles for energy system design, and suggests that good decisions can be made about the design and control of district energy systems without gathering building specific load data.

Finally, the value of having the option to adjust asset sizings for the design of a grid-scale energy park, and change energy storage technology choice, after uncertainty in storage performance has reduced is quantified. In contrast to building load, reducing uncertainty in storage technology performance by building a demonstrator system significantly reduces the cost of the energy park. Updating asset sizings after storage uncertainty is reduced is found to reduce total costs by 18% on average. While having the option to switch storage technology choice as well reduces costs by a further 13%, which is substantially greater than the cost of providing storage optionality.

The benefit that data and uncertainty reduction provide to decision making in energy systems is not obvious, even to those with expert knowledge of the systems. The results of this thesis show the importance of systematically studying how data can support decision making, and the limitations of the current static view of data and uncertainty in the energy systems field.

Table of contents

List of figures	xix
List of tables	xxiii
List of abbreviations	xxv
1 Introduction	1
1.1 Decarbonising an uncertain future	3
1.2 Existing literature	6
1.2.1 Uncertainty in energy systems	6
1.2.2 Understanding and accounting for the impact of uncertainty	8
1.2.3 Uncertainty reduction to improve decision making	16
1.2.4 Data collection requirements & costs	21
1.3 Thesis overview	24
1.3.1 Research gap and thesis aims	24
1.3.2 Thesis structure	25
1.3.3 Open data and models	26
1.3.4 Research outputs and contributions	26
2 Value of Information analysis	29
2.1 Introduction	30
2.2 Statistical decision theory	31
2.3 Bayesian decision analysis	37
2.4 The Value of Information	44
2.4.1 Perfect Information	44
2.4.2 Partial Perfect Information	46
2.4.3 Imperfect Information	47
2.4.4 Information theoretic Vol	50
2.5 Framework extensions	50

2.5.1 On-Policy Value of Information	51
2.5.2 Value of Optionality	53
2.6 Limitations of framework	54
Appendix 2.A Simple building energy system model for example problem	58
2.A.1 Energy model	58
2.A.2 Probabilistic model of uncertainties	61
2.A.3 A brief tirade on constraints in statistical optimisation	62
3 Applying Vol to energy systems	63
3.1 Introduction	64
3.2 ASHP maintenance scheduling	66
3.3 Office ventilation scheduling	70
3.4 GSHP borehole design	75
3.5 Conclusions	79
4 Forecasting data requirements for MPC	81
4.1 Introduction	83
4.1.1 Forecasting models for MPC	84
4.1.2 Impact of data on forecast accuracy and MPC performance	84
4.1.3 Research objectives & contributions	87
4.2 Experimental setup	89
4.2.1 Smart building control simulation framework	89
4.2.2 Forecasting task & performance evaluation	93
4.2.3 Building electricity usage data	94
4.2.4 Brief description of the prediction models	94
4.3 Results & Discussion	96
4.3.1 Baseline prediction accuracy comparison	96
4.3.2 Model generalisation	98
4.3.3 Data efficiency	101
4.3.4 Sensitivity of Model Predictive Control to forecast accuracy	107
4.4 Conclusions	109
Appendix 4.A Multi-building energy system asset specification	111
Appendix 4.B MPC planning horizon	112
Appendix 4.C Technical specification of prediction models and training procedures	113
4.C.1 Simple neural models	113
4.C.2 State-of-the-art machine learning models	113
Appendix 4.D Data similarity metric	114
4.D.1 Functional data analysis	114

4.D.2 Wasserstein similarity metric	115
Appendix 4.E Change-point analysis	117
5 Supporting solar-battery system sizing with load monitoring	119
5.1 Introduction	121
5.1.1 Quantifying the impacts of uncertainty on building energy system design	122
5.1.2 Research objectives & contributions	124
5.2 Case study district energy system	125
5.2.1 Case study data	125
5.2.2 Probabilistic models of uncertain building electrical load & measurement via monitoring	127
5.2.3 Stochastic Programming for system sizing	131
5.2.4 Evaluation of operational cost of system designs via simulation	135
5.2.5 Decision problem formulation	136
5.3 Results & Discussion	138
5.3.1 System design without load monitoring	138
5.3.2 Value of building load monitoring	140
5.3.3 Planning risk reduction	146
5.3.4 Practical importance & limitations of study	147
5.4 Conclusions	149
Appendix 5.A Case study system parameter specification	151
Appendix 5.B Details of prior system design	153
Appendix 5.C Investigation of numerical accuracy & computational requirements	156
5.C.1 Statistical accuracy	156
5.C.2 Cost estimation accuracy	157
5.C.3 Computational cost	160
6 Hedging storage technology selection for energy park design	161
6.1 Introduction	163
6.1.1 Storage technology performance uncertainty in energy park design .	164
6.1.2 Storage technology selection in energy park design	166
6.1.3 Research objectives & contributions	166
6.2 Experimental setup	168
6.2.1 Illustrative energy park system	168
6.2.2 System design under uncertainty	168
6.2.3 Probabilistic models of renewable generation & industrial load	170

6.2.4 Probabilistic models of storage technology performance & measurement	172
6.2.5 Stochastic Programming for system design	175
6.3 Results & Discussion	179
6.3.1 Prior design	179
6.3.2 Design updating without optionality & the <i>Value of Information</i>	179
6.3.3 Designing with optionality & the <i>Value of Optionality</i>	181
6.3.4 Sensitivity analysis	184
6.3.5 Use of multiple storage technologies	186
6.3.6 Risk-averse design	188
6.4 Conclusions	189
Appendix 6.A Parameter settings used in experiments	191
7 Conclusions	193
7.1 Limitations and further work	198
Appendix A Cambridge University Estates building energy usage dataset	201
References	205

List of figures

2.1	Decision tree representation of decision making under uncertainty	32
2.2	Decision tree representation of decision making with measurements	38
2.3	Example reduction in load uncertainty provided by metering data.	41
2.4	Influence diagram representation of example energy system design problem. .	42
2.5	Distribution of solar-battery system designs made with metering data.	43
2.6	Comparing distributions of possible system costs for designs made with and without perfect information.	45
2.7	Illustrative EVII vs. measurement cost curve	47
2.8	Comparison of initial design and designs made with reduced load uncertainty.	48
2.9	Illustration of energy flows in example building energy system.	58
3.1	Influence diagram representation of ASHP maintenance scheduling decision problem	68
3.2	Distribution of utilities achieved by optimal prior action, $N_m^* = 2$. Dashed line indicates mean of distribution	69
3.3	Distribution of true optimal maintenance frequency, i.e. best action once uncertainty has been revealed	69
3.4	Influence diagram representation of building ventilation scheduling decision problem	72
3.5	Distribution of utilities achieved by each ventilation rate under prior uncertainty	72
3.6	Distribution of utilities achieved by prior and pre-posterior decisions	72
3.7	Distributions of utilities for prior solution with varying assumed infection rate	72
3.8	Influence diagram representation of GSHP heat supply system design decision problem	77
3.9	Expected Value of Imperfect Information (EVII) and net economic benefit (EVII minus measurement cost) of imprecise ground thermal conductivity measurements	78
4.1	Graphical abstract of Chapter 4	82
4.2	Control loop for test district-level building energy system simulation.	89

4.3 Energy flow schematic for test multi-building energy system. (Icon credits: Symbolon)	90
4.4 Comparison of baseline prediction accuracy and operational performance of forecasting models.	97
4.5 Comparison of forecasting model generalisation. Violin plot with quartiles indicated by horizontal lines.	99
4.6 Correlation between Linear model generalisation and training dataset similarity metric values.	101
4.7 Average improvement in prediction accuracy over all prediction variables with years of data used for training.	102
4.8 Improvement in prediction accuracy of models trained using data from change-points onward.	103
4.9 Pearson correlation between data variables for Building 0.	105
4.10 Variation of prediction accuracy with no. of data features included in Linear model.	105
4.11 Variation of prediction accuracy with online update frequency.	106
4.12 Sensitivity of operational performance to synthetic forecast noise simulating prediction inaccuracy.	108
4.13 Variation of perfect forecast MPC operational performance and computation time with planning horizon, T	112
4.14 Schematic process for functional principal component analysis (fPCA)	116
4.15 Illustration of first two Phase (H) and Amplitude (V) fPCs. Solid black line shows mean function, $\mu(t)$. Dashed line (-) indicates the impact of a +ve coefficient and dotted line (.), a -ve coefficient.	116
4.16 Example fPC coefficient distributions for Buildings 38 (pink) and 49 (green)	116
4.17 Kernel density plot of V1 scores for Buildings 38 & 49. Wasserstein distance between two distributions is 0.05.	116
4.18 Similarity metric scores between training datasets for pairs of buildings.	117
4.19 Similarity metric scores between pairs of training, validation, test datasets for each building.	117
4.20 Change-points detected using the BEAST algorithm in load profiles of Buildings 26 & 44 (top row). Signal is decomposed into a seasonal component (second row), assumed harmonic, inferred change-points indicated by red vertical lines, and a trend component (fourth row), with green lines indicating change-points. Credible intervals of the estimated signals are given by grey envelopes around the individual components.	118
5.1 Graphical abstract of Chapter 5.	120

5.2	Example electricity load profiles, representing buildings with electrified heating.	126
5.3	Visualisation of prior probabilistic model of uncertain building load, with profiles sampled from distribution.	129
5.4	Visualisation of posterior probabilistic model of building load for example measurement.	130
5.5	Energy flow schematic for case study district energy system.	131
5.6	Methodology for computing Vol for case study district energy system design task.	137
5.7	Distribution of simulated total system costs and LCOEs for prior system design over sampled building load scenarios.	139
5.8	Total asset capacities in optimized system designs for each hypothesised load monitoring measurement. <i>Prior system design indicated with diamond.</i>	141
5.9	Correlation between total asset capacities in posterior designs and mean load over posterior sampled building load scenarios used for design (from scenario reduction).	141
5.10	Comparison of simulated total system cost distributions for prior and pre-posterior cases.	142
5.11	Comparison of simulated LCOE distributions for prior and pre-posterior cases.	143
5.12	Distribution over hypothesised measurements of standard deviation and range of total cost distribution of optimized system design for each hypothesised measurement (simulated for scenarios drawn from corresponding posterior distribution).	147
5.13	Illustration of operation of prior system across seasons.	154
5.14	Convergence of Monte Carlo estimates of expected utilities with number of samples from building load distribution.	156
5.15	Distribution of Stochastic Program objective value errors compared to simulated mean operating costs for posterior designs in base case.	159
6.1	Schematic of energy flows within energy park model	169
6.2	Two-stage decision model of energy park design	169
6.3	Examples of wind and solar normalized generation profiles for Rotterdam. .	171
6.4	Prior distribution of storage efficiency, and corresponding posterior for an example measurement value.	174
6.5	Final designs of energy park with NaS storage for each sampled measurement.	180
6.6	Distribution of final design expected costs over sampled measurements for each storage technology.	181
6.7	Final designs of energy park with storage optionality for each sampled measurement.	183

6.8 Reduction in average cost of final design provided by storage technology optionality, and selected technology, for each sampled measurement.	183
6.9 Operation of energy park with Li-ion and CAES storage, demonstrating different energy arbitrage time scales.	186
A.1 Summary of data availability across buildings in dataset.	204

List of tables

2.1	Deterministic design optimisation results.	35
2.2	Stochastic design optimisation results.	36
2.3	Summary of designs made with different levels of uncertainty reduction and associated Vol values.	49
2.4	Description of variables & parameters of Stochastic Program, Eq. 2.20. .	60
2.5	Parameter values for example building energy system model.	61
2.6	Distributions of uncertainties.	61
3.1	Sensitivity of prior solution and EVPI to office size	74
3.2	Sensitivity of prior solution and EVPI to infection prevalence in the population	74
3.3	Uncertainty and cost of ground thermal conductivity measurement tests .	78
4.1	Description of Linear Program model parameters.	92
4.2	Computation times for baseline models, trained on 8 years of data, and predicting for simulations of 2 years duration.	98
4.3	Specification of distributed energy assets in simulated multi-building energy system.	111
4.4	Technical specifications of complex, state-of-the-art models.	114
5.1	Probabilistic models of parameters describing uncertain building load, and their measurement via monitoring.	128
5.2	Description of Stochastic Program variables & parameters.	133
5.3	Expected system operating cost and Vol for separate uncertainty reduction in load parameters.	144
5.4	Parameter values for district energy system model used in experiments. .	152
5.5	Settings for sampling from probabilistic load model used in experiments.	152
5.6	Installed asset capacities for prior system design.	153
5.7	Breakdown of cost contributions over sampled scenario simulations for prior system design.	153

5.8 Errors in SP estimates of mean system operating costs for optimized designs (under-estimation +ve).	158
6.1 Probabilistic models of renewable generation and industrial load	171
6.2 Probabilistic models of uncertain storage technology performance parameters and their measurement.	174
6.3 Description of Stochastic Program variables & parameters.	177
6.4 Performance of initial energy park designs using different storage technologies.	179
6.5 Sensitivity of Value of Information and Value of Optionality to level of uncertainty reduction.	184
6.6 Sensitivity of Value of Information and Value of Optionality to electricity prices and carbon emissions	185
6.7 Comparison of energy park performance for different design cases.	187
6.8 Value of Information and Value of Optionality for risk-averse design cases .	188
6.9 Parameter values for energy system model used in experiments.	191
6.10 Settings for sampling from probabilistic models used in experiments.	192
A.1 Summary of variables available in dataset.	203

List of abbreviations

ASHP	Air Source Heat Pump
BEV	Battery Electric Vehicle
CAES	Compressed air energy storage
CCUS	Carbon Capture, Utilization, and Storage
COP	Coefficient of Performance
CVaR	Conditional Value-at-Risk
EV	Electric Vehicle
EVII	Expected Value of Imperfect Information
EVO	Expected Value of Optionality
EVP	Expected Value Problem
EVPI	Expected Value of Perfect Information
EVПPI	Expected Value of Partial Perfect Information
EVSI	Expected Value of Sample Information
GSA	Global Sensitivity Analysis
GSHP	Ground Source Heat Pump
HVAC	Heating, Ventilation, and Air Conditioning
LCOX	Levelized Cost of eXergy
Li-ion	Lithium ion battery
LP	Linear Programming
MC	Monte Carlo
MGA	Modelling to Generate Alternatives
MPC	Model Predictive Control
NaS	Sodium–Sulfur High Temperature battery
NZEB	Net Zero Energy Building
RBC	Rule Based Control
RL	Reinforcement Learning
RO	Robust Optimisation
SA	Sensitivity Analysis
SP	Stochastic Programming

TMY	Typical Meteorological Year
UA	Uncertainty Analysis
VoI	Value of Information
VoO	Value of Optionality
VRFB	Vanadium Redox Flow battery
VSS	Value of Stochastic Solution

1

Introduction

Few things have changed the world more than the harnessing of electricity to improve the lives of humans. However the harnessing of data over recent decades has begun to rival that impact. We are currently attempting to fundamentally redesign our electricity systems to remove the carbon-intensive infrastructure they were originally built around. To help achieve this, we are looking to data to improve our understanding of these complex energy systems, and to guide us to make better decisions. But data is itself a resource. It's costly. To use it effectively we need to understand what information it can give us, how it can improve our decision making, and ultimately when and where it's worthwhile collecting.

Decarbonisation is a truly enormous challenge (MacKay, 2016). The UK and EU have committed to getting to net zero carbon emissions by 2050 (Committee on Climate Change, 2020; European Union, 2021). There are lots of ways of slicing up the carbon pie. But from a scientific perspective, there are three broad areas to decarbonise: *industrial processes*, where chemists must figure out how to make the materials we need without producing CO₂ along the way; *agriculture*, where biologists need to find ways to grow enough food without releasing CO₂ and methane; and *energy*, where physicists and engineers have to produce

🔊 This introduction is quite long and wordy. You can listen to it [here](#) instead.

the electricity and heat that support modern life without using fossil fuels. This all needs to be done at a cost that is low enough that society is willing to make the transition, and without ever interrupting supply. Imagine rebuilding your tennis racket out of new materials while making sure you stay in the rally.

Within energy, many sectors such as heating and transportation plan to decarbonise by electrification (potentially using hydrogen to store and transport the energy). A key example of this is the shift towards using heat pumps for space heating (House of Commons, 2022; IEA, 2021). As a result, electricity demand is expected to increase by a factor of 3 to 4 by 2050 (ESO, 2023). Decarbonising this electricity will require vast amounts of renewable power generation to be integrated into the grid. It's estimated that around 100 GW of wind and 60 GW of solar generation will need to be constructed in the UK by 2050 (ESO, 2023). To support the new generation mix with mostly renewables, the grid will need to be redesigned to manage the variability of these new energy sources, and make sure electricity is available when it's needed. Adapting the electricity system to overcome this challenge of security of supply (Papadis and Tsatsaronis, 2020) will be costly, on top of the already hefty cost of building the renewables. It will require building new supporting infrastructure, primarily transmission lines (ESO, 2024) and bulk energy storage (Sepulveda et al., 2021), at scales rivalling the original construction of the power grids, and new methods for operating the grid to balance the system far more quickly and plan over horizons of months rather than days.

This redesign of the electricity system to match renewable supply to consumer demand provides a traditional, top-down view of energy provision. However, in recent years, a new perspective has emerged. This perspective considers a bottom-up approach, where consumers play an active role in the energy system. By shifting when they use energy, consumers can help match their demand to the available renewable supply on the grid, reducing the need for supporting infrastructure. This demand flexibility can help both reduce the cost of decarbonisation and speed up the process.

As buildings in the EU currently use around 40% of energy, and are responsible for about 35% of carbon emissions (EEA, 2023; IEA, 2023), redesigning building energy systems could make a large contribution towards decarbonisation targets. There are many ways to reduce the carbon emissions of buildings, such as improving their energy efficiency through fabric and equipment upgrades, and incentivising adaptations in user behaviour to unlock demand flexibility. One of the most effective and cost efficient approaches is to install local renewable generation and energy storage (Amini Toosi et al., 2022; Leibowicz et al., 2018; O'Shaughnessey and Shah, 2021). Combining this with smart control of the building energy system can further reduce operational costs and carbon emissions

(Kathirgamanathan et al., 2021; Nweye et al., 2023). The benefits can be increased again if the operation and/or design of multiple nearby building energy systems are coordinated, forming a “district energy system” (Pickering, 2019; Vazquez-Canteli, Henze, et al., 2020). This allows much of the energy demand to be provided locally, and can significantly reduce the impact those buildings have on the grid, for example decreasing peak load and so the need for transmission infrastructure (Niveditha and Rajan Singaravel, 2022).

There are roughly 25 million homes (DESNZ, 2024b; X. Li et al., 2022) and 1.5 million commercial buildings (DESNZ, 2024c) in the UK that need to be retrofitted. On top of this, several million more buildings will be constructed before 2050. Designing energy systems for these buildings presents both an enormous challenge and opportunity.

1.1 Decarbonising an uncertain future

So, to reach net zero we need to design, build, and effectively operate new energy systems across scales from the individual building to the national level grid. When planning these systems, we must confront the inherent uncertainty of the task. The systems will operate in a future which we can only imperfectly predict. When designing, the weather patterns determining renewable generation and heat demand, the cost of materials needed to build the infrastructure, the efficiencies different technologies will achieve, and the energy demand that needs to be provided are all uncertain. When operating, the price and availability of power, when consumers will want to use energy, how much energy they will need to heat their homes that day, how those consumers will respond to price signals, the condition of equipment in the system, and the current state of the system are all uncertain.

Many of these uncertainties are exacerbated by the system adaptations needed to decarbonise. For example, the more renewables are used for power generation, the greater the dependence of electricity price on uncertain weather conditions. Similarly, as more heat pumps are installed for space heating, electricity demand will become increasingly dependent on those same uncertain weather conditions.

From a planning perspective, uncertainty is problematic because it introduces the risk of high cost outcomes which occur when conditions end up being unfavourable. Mitigating this risk requires making decisions that hedge against the uncertainties, but doing so increases the costs in favourable scenarios. So uncertainties induce costs. Managing the uncertainties in energy systems is therefore crucial to reducing the cost of decarbonising energy, and making the transition affordable for society.

Some uncertainties are intrinsic and are a result of inherent randomness in a system. These cannot be reduced. For example, how much energy the solar panels in the UK

will produce next year will always remain uncertain. We call these aleatoric uncertainties (Pelz et al., 2021). In some cases, other uncertainties *can* be reduced by gathering data and improving our knowledge about the system. For instance, by monitoring how many people use a building and gathering their temperature preferences, we can reduce uncertainty in how much electricity will be needed to heat the building tomorrow. We call these epistemic uncertainties (Pelz et al., 2021). Most often uncertainties are a mix of both types¹, meaning we can gather data to improve our knowledge, but that there is a limit to how much uncertainty can be reduced.

Aleatoric uncertainties can only be addressed by improving planning methods to better handle the uncertainty. Epistemic uncertainties however can also be addressed by collecting data. This leads us to a data-driven paradigm for energy system planning, and raises a host of important questions about the data we want to collect:

- How much does collecting data reduce the uncertainties?
- Which data points lead to greater uncertainty reduction than others?
- Do some data points actually contain the same information and so have redundancy?
- How much will reducing uncertainty actually improve planning?
- How much will it cost to collect different data?
- Which data is best to collect?
- And how much of it do we need?
- Is it actually worth collecting any new data at all?

Data-driven methods are becoming ubiquitous in both research (Girolami, 2020; Ying Sun et al., 2020a) and engineering practice (ASHRAE, 2024). They are a key part of developing functional smart energy systems (Huiliang Zhang et al., 2021), and achieving the potential benefits of Digital Twinning (Wagg et al., 2025).

However, whilst smart buildings and Digital Twins have been researched and discussed with great emphasis for many years, there is still limited consensus on what the scope of these systems should be (Bortolini et al., 2022; Semeraro et al., 2021). For instance, what data they should collect. While smart energy systems are continuing to gain traction, and regulation has aimed to accelerate their adoption (Clements, 2020), they are still far

¹ If you think hard enough about most aleatoric uncertainties you will probably come to the conclusion that better modeling might be able to reduce uncertainty. Whether this modeling is practically achievable, and how much it would actually reduce uncertainty is another matter. *With a complete physics model of the solar system, weather could be perfectly predicted. But in practice we are much more likely to accept weather as inherently uncertain.*

from the norm. In the UK, only 51% of electricity and gas meters are currently ‘smart’ (able to collect hourly energy usage data), despite rollout beginning in 2012, and expected to cost £19.4bn in total (DESNZ, 2023f). Little information is available on the extent of adoption of other smart building technologies. Only recently was a standard for measuring the ‘smartness’ of buildings analogous to the EPC rating proposed by the Centre for Net Zero (CNZ, 2025).

Smartness, and the data it provides to support planning, will be crucial for effectively managing the uncertainty of decarbonised energy systems. However, designing, deploying, and maintaining smart systems is both highly complex and costly.

Data usage and data requirements have been studied for several decades in other fields, such as medicine, chemistry, and information theory. This has led to the development of several methods for quantifying when and where data should be collected, such as Bayesian optimization (Shahriari et al., 2016), Bayesian experimental design (Rainforth et al., 2024), and Value of Information analysis (Vol) (Keisler et al., 2014), the main method studied in this thesis. However in the context of energy systems, data requirements and the importance of uncertainty reduction to support decision making have so far received limited attention. On top of this, while smart energy systems are widely discussed, the costs of building and maintaining this smart infrastructure, and how the data they collect should be used, often go unmentioned.

To make effective decisions about the use of data collection to manage the uncertainties of decarbonised energy systems, and support planning during design and operation, we need to understand the benefits that data provides for improving decision making. To be able to justify, rationalise, and prioritise expenditure on data collection, the value of data for supporting decision making must be quantified. With this, we can determine which data is most beneficial to collect, how much data should be collected and where, and ultimately whether the benefits of collecting more data are worth its cost.

1.2 Existing literature

1.2.1 Uncertainty in energy systems

Uncertainties are everywhere in energy systems. They affect almost every aspect of an energy system, and appear in a diverse range of forms across the different system scales. The planning horizon being considered also influences the way in which uncertainties affect the energy system. For example, when controlling a domestic solar-battery system, uncertainty in temperature over the next few days leads to uncertainty in the heat demand of the building, and so the operating cost of the system over that short planning horizon. But, when sizing that same system, uncertainty in the minimum and maximum temperatures over the year becomes important, as the system should be designed to be able to meet the peak heating and cooling demands of the building.

Properly modelling the uncertainties affecting an energy system is crucial to being able to correctly determine the impacts they have on the system. However, as the appropriate model of uncertainty depends highly on the subtleties of the energy system being modelled (such as the scale of the system studied, the scope of the research question, or the specifics of the energy model used), it is very difficult to provide firm guidelines for the form that probabilistic models of uncertainties should take, or even the broader methodology of modelling uncertainties (Chong, Xu, et al., 2015; Jane Fennell et al., 2019; Tian, 2013). As such, in the literature uncertainty modelling is done on a case-by-case basis (Chong, Xu, et al., 2015; Prataviera et al., 2022), often with limited information provided about how the uncertainties are modelled and the choices made during modelling (Chengshan Wang et al., 2016).

As there are numerous uncertainties in any energy system, most studies select just a few to consider during their analysis (Gabrielli et al., 2019; Tian, Heo, et al., 2018; Yue et al., 2018). In some cases pre-screening methods are used to identify the most influential uncertainties, which are then studied in a more detailed uncertainty analysis (Mavromatidis et al., 2018b). Exhaustively studying all uncertainties is impractical, and in most cases the uncertainties considered are in fact the result of a larger set of uncertainties interacting with some physics in an underlying, higher resolution energy model.

For example, in this thesis energy demand and the resulting uncertainty are treated as exogenous, i.e. it's assumed that the energy systems studied are required to meet a given demand that is unknown/uncertain but not influenced by how the energy systems are designed or operated. However, in reality energy demand in buildings is predominantly caused by human behaviours, which are both inherently uncertain, and is influenced by

uncertain environmental conditions such as weather, and uncertain macro-scale energy system conditions such as electricity price. Detailed modelling of the way occupants interact with buildings and use energy, for example to maintain their thermal comfort, and how this interaction can be managed to assist decarbonisation, are active areas of research (M. Hu and Xiao, 2020; O'Neill and F. Niu, 2017; Tröndle and Choudhary, 2017; Ward, 2021). However they're beyond the scope of this thesis.

Selecting which uncertainties to consider when modelling energy systems is a crucial step. Missing an influential contribution to uncertainty can lead to underestimation, and because when looking at uncertainty reduction, the interaction of uncertainties can have a significant effect on results, as is seen in Chapter 5.

Some uncertainties are common to energy systems across all scales, for example outdoor temperature and energy demand. But the impacts they have on the energy systems depend highly on the scale and purpose of the system. This has led to the development of standard datasets which can be used for uncertainty analysis by integrating the data with the specific energy model being studied. A notable example of this is the range of [Typical Meteorological Year \(TMY\)](#) datasets (Crawley and Lawrie, 2015; Wu et al., 2023) created for building energy simulation, including extensions to consider future climate uncertainties (Bravo Dias et al., 2020). However, by nature of their generality, they are not well suited to every use case (Rady et al., 2025). Another challenge where uncertainties are common is ensuring that the representations of uncertainty are consistent across system scales. For example, bottom-up models of uncertain energy usage are calibrated to match the variability of known aggregate demand patterns (Chong and Menberg, 2018). Similarly, consistency of uncertainty representations across temporal scales is important. For example, ensuring models of hourly energy demand are consistent with distributions of daily and annual values (Ward, 2021).

Some uncertainties can be well represented by simple statistical models. For example, the energy storage technology performance metrics considered in Chapter 6 are modelled using independent Gaussian distributions. However, many uncertainties in energy systems have complex underlying patterns, and require more sophisticated statistical models. A key example of this is the complexity of patterns over varying time scales that occur in energy and weather time series. In this case, substantial modelling effort is needed to properly represent the statistics of this functional data (Ward, 2021). The complexity of uncertainties also affects how they need to be represented in energy system models, and analysed, to correctly determine the impact they have on the system.

Ultimately, some data is needed to develop probabilistic models, and data availability determines the complexity of model that can be created. Gathering more data from an energy system allows for a more detailed and precise probabilistic model to be developed.

1.2.2 Understanding and accounting for the impact of uncertainty

The importance of understanding the impact that uncertainties have on the behaviour of energy systems, and how energy systems should be designed and operated to manage this uncertainty, has been widely discussed in the literature (DeCarolis et al., 2017; Khezri et al., 2022; Xiwang Li and Wen, 2014; Pfenninger, Hawkes, et al., 2014; Prataviera et al., 2022; Rysanek and Choudhary, 2013). The study of uncertainty is now commonplace in the energy systems field. Many review papers have covered the different techniques proposed for assessing uncertainties and their impacts (Alonso-Traveset et al., 2022; Fennell et al., 2025; Liu and Y. Wang, 2020; Macdonald and Strachan, 2001; Majidi et al., 2019; Mavromatis et al., 2018a; Rivalin et al., 2018; Tian, Heo, et al., 2018; Yue et al., 2018). And these methods have been applied to the design and control of energy systems across all scales, from individual buildings with local generation and storage (Coppitters et al., 2021), to national scale power grids (Neumann and Brown, 2021). However, despite the widespread use of uncertainty analysis, and the demonstration of its importance for properly understanding the performance of energy systems, some studies still don't include uncertainties in their modelling (Fiorentini et al., 2023; Jinshi Wang et al., 2022; Yue et al., 2018).

Broadly speaking, there are two main approaches² to studying the impact uncertainties have on energy system models: evaluation based methods, where uncertainties are treated as inputs to a given model, and representation based methods, where uncertainties are represented within the model itself.

Evaluation based methods

Evaluation based methods start with a given energy system model that is to be studied. Uncertainties in the energy system are considered to be external to the model, and are treated as model inputs. The effect of these uncertainties is studied by repeatedly sampling values of the uncertain parameters, and evaluating the model with this input data. This builds up a picture of how the model responds over the distribution of the uncertainties, and various metrics can be computed to quantify the impact of the uncertainties on the model.

The key benefit of this approach is that it is very simple, and provides clear, interpretable results. As a result, it can be applied to complex energy system models³, and work with complex statistical representations of uncertainties. Because the statistical sampling, model evaluation, and analysis of results are separated, these techniques can be used in

² This thesis focuses on techniques for studying the impact that uncertainties have on energy system performance, rather than methods for characterising and modelling uncertainties, which is a separate but dependent area of study.

³ In fact they can be used with any arbitrary simulator.

a plug-and-play fashion⁴, making them very flexible, easy to work with, and approachable to non-expert users.

The main limitation of evaluation based methods is that, in separating the statistics and energy model, they can only study one instance of the energy system (e.g. a configuration or design) at once. So, while they can investigate how uncertainties affect the performance of that energy system design, they cannot answer questions about how the uncertainties impact decision making in the system, as those decisions are wrapped up in the simulator model being tested. In order to test how different decisions impact performance, the method must be repeated for each decision considered (using the corresponding simulator). This is highly computationally expensive. Though these stochastic simulations can be easily parallelised, this approach quickly becomes computationally infeasible for practical energy system models, or can only be used for very simple studies comparing a few different design choices.

Many different evaluation based methods have been used to investigate how uncertainties affect energy system models, including:

- Scenario analysis & probabilistic methods
- **Sensitivity Analysis (SA)** techniques
 - Local sensitivity analysis
 - One-at-a-Time (OAT)
 - Derivative based methods
 - **Global Sensitivity Analysis (GSA)**
 - Morris method
 - Sobol indices
 - Random forest feature selection
- Interval analysis
- Probabilistic methods
 - **Uncertainty Analysis (UA)**
 - **Monte Carlo (MC)** simulation
 - Statistical scenario analysis (e.g. scenario trees)
- **Modelling to Generate Alternatives (MGA)**

Review papers (DeCarolis et al., 2017; Fennell et al., 2025; Majidi et al., 2019; Tian, Heo, et al., 2018; Yue et al., 2018) provide detail on how the different methods work, and how they have been applied to the study of energy systems.

⁴ This feature also makes evaluation based techniques well suited to acceleration by compute parallelisation.

These methods have been used to study the impact of uncertainties across the full range of energy system scales, from individual buildings to inter-national power grids.

(Gang et al., 2015) performs an [Uncertainty Analysis](#) to investigate how a set of nine operational uncertainties in a high-rise building, including outdoor temperature and indoor set-points, affect the energy usage performance of a set of cooling system designs. For each chiller configuration considered, the input uncertainties are propagated through the energy model to produce distributions of peak cooling load and annual energy consumption. These distributions are then used to compare the designs, and select the best performing configuration, trading-off the risk of not being able to satisfy cooling load against the expected energy consumption in a holistic manner. Producing distributions of cooling performance provides a more complete picture of how the cooling systems will perform during actual operation, and so this study provides an improvement on the conventional, deterministic performance calculations used for [HVAC](#) sizing. However, as the uncertainties are not treated within the energy model, the uncertainty analysis can only be applied to the system designs initially chosen by the designer, and the method is unable to determine the true optimised design which provides the best trade-off between average cost and risk of not satisfying cooling load, as would be possible with a representation based method.

In (Mavromatidis et al., 2018b), [Global Sensitivity Analysis](#) is used to identify the uncertainties which have the greatest contribution to the variability of operating cost for a hypothetical urban district energy system located in Zurich, Switzerland. Deterministic design optimisations are performed for a set manually selected design scenarios, and a basic scenario analysis is performed comparing the cases. The uncertainty in the cost of operating the district system in each design case is then investigated, considering over 30 uncertainties, in energy carrier prices, emissions factors, investment costs, equipment efficiencies and losses, energy demand and generation profiles. Both the Morris method and Sobol indices are used to determine which of these uncertainties are responsible for the greatest portion of the operating cost variability. Uncertainty in energy carrier prices and the ratio of building energy demand to local solar generation are found to be responsible for the majority of the output variability. However, due to the limitations of these [GSA](#) methods, the analysis ends here. A large number of uncertainties are considered, and it is possible to identify a subset of those which do not make a significant contribution to the output variability, and so can be treated as deterministic without affecting the accuracy of the output distribution⁵. However, for the uncertainties which are identified as having large contributions to output variability, the results cannot inform the designer about whether these uncertainties are actionable, if trying to reduce them is worthwhile, or whether the

⁵ Though given that for evaluation based methods including more uncertainties does not increase computational cost, it's not clear why this would be beneficial if the designer already has access to the probabilistic models of these uncertainties, as is required to screen them out initially.

system design could be adjusted to better manage them. It can only indicate that modelling these uncertainties is important to properly capture the distribution of operating costs for the specific system design investigated.

In the context of large-scale power system design, (Brown, Schlachtberger, et al., 2018) uses scenario analysis to investigate the effect that different outcomes of uncertain future sectoral energy demands, energy infrastructure costs, and the availability of energy technologies such as long-duration storage have on the minimal cost design of the future European grid. A range of scenarios representing hypothesised future energy landscape conditions in Europe are proposed, with parameter values for each uncertainty suggested holistically based on the type change represented by the scenario. The energy planning model PyPSA-EUR model (Hörsch et al., 2018) is then run for each scenario, and the resulting energy system designs and their corresponding costs are compared. The study highlights the effect of particular parameter value changes on how decarbonised energy should be best provided, and the cost of doing so, by comparing pairs of scenarios, for instance allowing BEVs to participate in energy trading for grid balancing . It also attempts to pick out broader trends in the system designs across the scenarios to make recommendations on power system adaptations that are likely to be required and so should be planned for. Scenario analysis is the traditional method for handling uncertainties in energy system planning, and is still widely used in the academic literature, and even more so in industrial practice. The type of study performed in (Brown, Schlachtberger, et al., 2018) is a typical use of the scenario analysis methodology. It also clearly highlights its limitations for studying uncertainties. Scenario analysis is not (in its usual form) a statistical methodology. Specifying the parameter values used in the scenarios relies on expert judgement and is difficult to justify. Further, as there is no probability associated with each scenario, it is not clear how the results should be interpreted, and whether the recommendations from a particular scenario are likely to be relevant. Scenario analysis can be a useful tool to identify the design space that should be considered, but it is not suited to providing specific guidance on how to design energy systems. Nonetheless, it can be useful in situations where designers wish to look in detail at particular scenarios to observe and compare the mechanisms in the energy system model, and when limited statistical information is available, and so guesses at broad scenarios is the best available option. Future energy scenarios are a good example of this, for example those published by National Grid (ESO, 2023).

Modelling to Generate Alternatives provides a different approach to exploring the design space of energy systems. It does so by performing an initial deterministic design, and then exploring the space of designs which have model output (e.g. cost and carbon emissions) close to those of the initial design. This exploration can be done by sampling from the

uncertainties and repeating the design process with an updated objective function, or by considering search directions that correspond to particular scenarios of interest, e.g. the design with the smallest transmission expansion. (Pedersen et al., 2021) and (Neumann and Brown, 2021) apply this methodology to the same model of the European power system (Hörsch et al., 2018). The advantage of MGA is that it is able to demonstrate the impact that uncertainties have on the optimal design of an energy system, rather than just the cost/performance of a given design. However it suffers from a similar limitation to scenario analysis in that, while it shows the designer how the uncertainties affect the optimal system design, it doesn't provide any guidance on how the design should be adjusted to best manage those uncertainties.

Representation based methods

Representation based methods build energy system models which include representations of the uncertainties in the system. As the uncertainties are incorporated into the model, when the model is evaluated it provides results which account for those uncertainties, e.g. by returning the mean system performance and some measure of the uncertainty in that performance.

The advantage of this approach is that it allows the impact of uncertainties to be included in the decision making processes of the model. For example, an evaluation based approach would apply an optimization model to design energy systems that minimise operating costs for a number of scenarios sampled from the distribution of future operating conditions, and then use the average of those designs, claiming some 'robustness' against the uncertainties. Whereas, if those uncertainties are accounted for within the optimization model, the representation based approach can determine the system design that minimises the average operating cost of the system, or whichever aspect of the operating cost distribution the system operator is interested in. In this way, representation based methods are able to account for uncertainties during decision making, and so address more complex questions about the impact uncertainties have on energy systems.

However, achieving this more detailed insight requires more complex models. There are two aspects to this. Firstly, developing energy system models which include uncertainty representations is a more complicated task, requiring greater modelling effort, and understanding of both the physics of the energy system and the statistics of the uncertainties. Further, the models are often required to adhere to a particular mathematical structure (e.g. convexity) to enable the handling of uncertainty representations, which is both challenging and usually requires simplifying assumptions to be made, which limits the flexibility of the energy model and reduces its accuracy compared to a simulation based approach. As a result, these models are less approachable for non-expert users, and must

be developed for each energy system application. The second complexity is that these models are typically far more computationally expensive than simulations (for a given level of model detail). Further, they often have super-linear scaling, making large-scale systems intractable without substantial simplifications, limiting the applicability of these models.

Therefore, while representation based methods are necessary to properly understand how uncertainties affect decision making in energy systems, they can only be used to study small or simplified models.

The representation based methods used in the energy systems literature include:

- Stochastic Programming (SP)
- Chance/risk constrained optimisation
- Interval programming
- Robust Optimisation (RO)
- Fuzzy optimisation
- Information gap decision theory

The following review papers provide detail on how these methods incorporate representations of uncertainty their models, and how they have been used to study the effect of those uncertainties on decision making in energy systems; (DeCarolis et al., 2017; Liu and Y. Wang, 2020; Majidi et al., 2019; Yue et al., 2018).

Representation based methods have also been applied to energy systems across all scales. For example, (Coppitters et al., 2021) studies the design of a domestic building energy system which contains solar generation, an air-source heat pump, and both battery and thermal storage. Surrogate-assisted robust optimisation is used to find the pareto-front of designs that minimise the mean and standard deviation of the [Levelized Cost of eXergy \(LCOX\)](#) provided by the system under a set of 30 uncertainties including solar generation, energy prices, energy demands, and technical characteristics of the storage units. The purpose of this analysis is to inform the system designer about how the specification of the energy system can be adjusted to manage the uncertainties, and inform them about the trade-off between the mean cost and the risk of high costs. The results show that roughly doubling the asset capacities in the energy system can reduce the variability in exergy costs by around a third, at the expense of increasing the mean cost by around 20-25%. So, designing buildings to have greater energy independence can significantly reduce the uncertainty in the cost of running the energy system by reducing its reliance on uncertain electricity prices, but comes at a somewhat higher average cost.

(Yamamoto et al., 2024) investigates the use of robust optimisation in the context of building operation. It uses a robust **Model Predictive Control (MPC)** approach to schedule batteries and heat pumps in an apartment building with local solar generation. The controller aims to minimise the cost of supplying the building's energy demands, while ensuring that these energy demands can always be met, given that both the energy demand and solar generation in the building are uncertain over the planning horizon. Achieving this kind of security of supply robustness is relatively straightforward for a building controller, so this study considers how the operational cost achieved by the controller can be reduced while maintaining this robustness.

As these two papers show, managing the uncertainty in a building energy system comes at a cost. District energy systems have received a large amount of research interest as they present an opportunity to reduce the impacts of building level issues, such as building load variability/uncertainty, by aggregating multiple building systems to enable cooperation and resource sharing. This can provide substantial benefits for both the design and operation of the buildings. (Pickering and Choudhary, 2019) and (Mohammadi et al., 2022) investigate similar district energy system design tasks, but take different modelling approaches. Both studies consider the sizing of solar PV and energy storage for buildings within a district energy system, and aim to minimise the average total cost of the system, including both capital and operational costs. (Pickering and Choudhary, 2019) models the energy system via a **Linear Programming**, and uses **Stochastic Programming** for design, accounting for uncertainty in the building loads. Whereas (Mohammadi et al., 2022) takes a simulation based approach, using a genetic algorithm for the design optimisation, and considers uncertainty in solar generation, and the price and availability of grid electricity. Both approaches are able to account for the effects of the uncertainties on the operation of the district and determine designs that reduce the average cost of supplying the buildings' energy needs. The genetic algorithm approach provides greater flexibility in the behaviours of the district energy system that can be modelled and requires fewer simplifying assumptions, whereas the **LP** approach provides guarantees that the best possible design has been found.

(Min et al., 2018) considers a similar issue of ensuring security of supply to (Yamamoto et al., 2024), but in the context of long-term planning for a large-scale power system. In this case the focus is on how the uncertainty in renewable generation must be managed to ensure that sufficient energy is available to meet demands when decarbonising the Korean electricity system. Chance-constrained programming is used to determine the supply mix that minimises the total cost of providing electricity whilst having sufficiently low probability of being unable to meet demand, called the loss-of-load probability. The results show that, due to the uncertainty in renewable generation, increasing the strictness of the

security of supply requirements requires more renewable generation capacity to ensure system reliability, and so increases the cost of the electricity system. Therefore setting the loss-of-load probability requirement allows the system designer to trade-off the risks caused by generation uncertainty with the cost of mitigating them.

The importance of representation based methods is that they allow the management of uncertainty to be incorporated into the decision making process. This allows energy system designs or control strategies that best mitigate uncertainty in a particular desired way to be identified, rather than simply guessing at good designs and observing how uncertainty affects their performance ex-post using an evaluation based method. Adjusting the tolerance to the impacts of uncertainties allows designers to investigate how energy systems should be adapted to mitigate those impacts, and the costs of doing so. For example, installing more battery capacity in a building energy system can reduce the risk of very high energy costs at the expense of increasing the average cost of the system.

The differences between representation based methods are in how they model the energy system, how they represent uncertainties within that model, and how they are able to manage the effects of those uncertainties in the decision making process. As a result, the most appropriate method for a given application depends on the characteristics/behaviours of the energy system and its uncertainties that are most important for properly modelling it. For instance, in a hospital energy system where security of supply is critical, the ability to observe the full distribution of outcomes caused by the uncertainties and ensure worst-case scenarios can be managed is extremely important, and so a detailed simulation based approach is likely required even if it is not capable of much cost optimisation.

Combining both approaches

Evaluation and representation based methods are able to answer important questions about how uncertainties affect the performance of an energy system, and how those uncertainties should be managed during decision making, respectively. The two types of methods are compatible and could be used together to provide a more complete investigation of how uncertainties impact an energy system. An example of this would be the use of an evaluation based method to perform an initial screening of uncertainties to identify which do not have a significant influence on the energy system, followed by the use of a representation based method to perform an optimisation based design for the system considering the reduced set of uncertainties, then finally a more detailed uncertainty analysis of that design using an evaluation based method to investigate the uncertainty in its performance. The first half of this combined methodology is suggested during discussion in (Mavromatidis et al., 2018b).

1.2.3 Uncertainty reduction to improve decision making

The purpose of collecting data is to reduce uncertainty⁶. So, to understand how data collection improves decision making, we need to understand the effect that reducing uncertainty has on the decision making process. All of the methods for studying uncertainty and their applications to energy systems discussed in the previous section consider only one particular state of uncertainty, and inform us of the effects that uncertainty has on the energy system, and how it should be managed during decision making.

A few existing studies have performed analysis that provides a step towards investigating the impact of uncertainty reduction on energy systems decision making.

(Yongjun Sun, 2015) studies the optimal sizing of HVAC, solar and wind generation, and battery storage capacities in a simulated [Net Zero Energy Building \(NZEB\)](#). A set of 10 uncertain parameters are considered, related to the properties of the building, its thermal performance, and the efficiency of generation. These include the building's wall thickness and infiltration rate, the indoor temperature set-point, the [COP](#) of the [HVAC](#) system, and the solar PV efficiency. A local sensitivity analysis is performed to determine how changes in these parameters affect the optimal capacity sizings of the energy system, as well as the overall system cost. Influence coefficients for the effect each uncertain parameter has on the sizing of each system component and the cost are then computed using [GSA](#). It is found that the parameters related to the thermal performance of the building (internal thermal gain intensity, indoor temperature set-point, and [HVAC COP](#)) have the greatest effect on the optimal system design and total cost.

From this analysis we could conclude that measuring these parameters would have the greatest effect on the energy system design that would be chosen, and so this suggests that reducing uncertainty in these parameters might be important. However, this study only performs deterministic optimisation. So, there is no initial stochastic solution, which accounts for and manages the uncertainties in the design, to compare the changes in components sizings and costs to. The sensitivity analyses performed are also deterministic, with changes in parameter values taken from feasible ranges within consideration of whether those values are likely. Before collecting data it is not possible to know what values will be measured, so properly quantifying the effect data collection has on decision making inherently requires a statistical approach. Therefore the results of this study have significant limitations. They show how the optimal design and cost of the [NZEB](#) would change if the

⁶ This might seem odd and slightly simplistic to some. After all there are lots of complicated issues we may need to address by collecting data, such as ensuring adequate data coverage and representivity for training prediction models, measuring important states in real time monitoring systems, and so on. But these issues *can* all be viewed from the perspective of uncertainty reduction. Whenever we have the ability to estimate/predict something, data collection can be seen as a way of reducing the uncertainty in that estimate. However, doing the probabilistic modelling is not always possible or especially informative.

parameters were measured to have certain values, but as there is no initial stochastic design to compare against, the effect of uncertainty *reduction* cannot be seen. Further, as the analysis is not statistical, it is not possible to see whether significant changes in the design and cost would be likely to occur. However, the results do demonstrate the relative sensitivity of the energy system design to the different uncertain parameters, and importantly can identify parameters which do not have a significant effect on the optimal design and cost, and so can be treated as deterministic without affecting decision making. Also, all uncertain parameters considered are possible to measure, which is not the case in many other studies.

A similar sensitivity analysis is performed in (Sanajaoba Singh and Fernandez, 2018) when investigating the sizing of solar, wind, battery storage, and AC-DC converter systems for an isolated district energy system in rural India. Local sensitivity analysis is used to determine the changes in capacity sizings and total cost of energy provision caused by changes in the capital cost of the system components, and the annual average solar and wind generation.

(Mavromatidis et al., 2018b) addresses one of the limitations of (Yongjun Sun, 2015) by assigning probabilities to the uncertain parameters it considers and sampling values from their distributions. It optimises the sizing of boilers, heat pumps, thermal storage, and solar generation in an urban district energy system. This optimisation is performed for a large set of samples from 30 uncertain parameters, including energy prices, investment costs, energy demand, and generation profiles, to minimise the total cost (investment plus operation) given each set of sampled values. Distributions of the optimised capacities of the components in the system are visualised to demonstrate how the system design changes when the uncertainties are realised. However, a distribution of the resulting total costs is not provided. Instead Sobol indices are computed to quantify the relative effect of each uncertain parameter on the cost of the system. Additionally, no stochastic optimisation is performed to determine the decision that would be made in the initial state of uncertainty that could be compared against. So while this study does determine the distribution of decisions (system designs) that would be made if uncertainty were removed from the problem, there is nothing to compare these reduced uncertainty decisions to.

A stochastic optimisation which provides the design decision made in the initial state of uncertainty is performed in (Pickering and Choudhary, 2019). Scenario Optimisation is used to determine the sizings of gas boilers, solar thermal & PV panels, and electrical & thermal storage in a district energy system that minimise the average total cost of meeting the district's energy demands, considering uncertainty in the energy demand time series of the buildings. To do so, 500 samples are taken from the building energy demand distributions, and a deterministic optimisation is performed for each sample. Scenario

reduction is then used to determine 16 representative scenarios to use in the Scenario Optimisation. The aim of the paper is to investigate the robustness of the stochastic design to “out of sample tests”, i.e. samples not included in the scenario optimisation. But while performing the scenario reduction, the analysis computes how the optimal design and cost changes with the realisations of the uncertain building energy demands. However, the paper does not report these results. It does compare the distributions of investment and operation costs for the optimal designs without uncertainty to the costs of the stochastic design. But as it does not combine them and compare total costs, it does not quantify the reduction in cost that would result from removing uncertainty in the problem to improve decision making.

While these studies did not explicitly aim to investigate the effect of uncertainty reduction on decision making, they do consider similar questions and perform some of the necessary analysis steps. In the case of (Yongjun Sun, 2015), (Sanajaoba Singh and Fernandez, 2018), and (Mavromatidis et al., 2018b), the focus is on the ‘sensitivity’ of the energy system designs to the uncertain parameters. This provides some broad indication about which uncertainties are important to the design process and so may be beneficial to reduce by taking measurements. However, ultimately decision makers are not concerned with what action they end up taking, but with their objective and how well they can perform the decision task (for example how much the cost of an energy system can be reduced). So it is unclear how these results showing the changes in optimal system design should be interpreted and actioned. They cannot inform the decision maker about whether reducing the uncertainties will significantly improve their decision making, only that some uncertainties do not affect it, and that others will change the actions they take. While these studies give some indication of the importance of uncertainties for decision making, none of them quantify the improvement in decision making that reducing uncertainty would provide.

Value of Information analysis (Vol**)**

Understanding the benefit of collecting data to reduce uncertainty and improve decision making is extremely important in many contexts. This research topic has been studied extensively in several different fields, and many approaches for assessing the benefits of uncertainty reduction have been developed. The most relevant method to decision making in energy systems is **Value of Information** analysis (**Vol**). This is because it is built on Statistical Decision Theory and Stochastic Optimization, and so aligns with the design optimisation approach commonly taken in the study of energy systems. **Vol** provides a framework for rigorously quantifying the improvement in decision making provided by a given reduction in uncertainty. In situations where optimisation is used for decision making and Bayesian probabilistic models of uncertainty reduction are available, **Vol** provides a

full mathematical treatment of how uncertainty reduction impacts decision making. **Vol** is the main methodology used in this thesis. Chapter 2 explains the methodology and its background in detail.

A large number of studies have applied **Vol** to investigate uncertainty reduction in fields such as environmental science, medicine, agriculture, and economics (Keisler et al., 2014). For example in medicine, **Vol** has been used to determine the sample size for randomised clinical trials that provides the best trade-off between reducing uncertainty in the efficacy of treatments and the cost of performing trials (Willan and Pinto, 2005), and to decide whether further research is required before a new medical technology is adopted (Tuffaha et al., 2014).

Within engineering, **Vol** has been used to study the benefit of collecting data to reduce uncertainty and improve decision making in contexts such as maintenance scheduling for buildings (Grussing, 2018) and wind farms (Myklebust et al., 2020), construction project planning (Esnaashary Esfahani et al., 2020), sensor placement (Malings and Pozzi, 2016), and structural health monitoring (Di Francesco, Chryssanthopoulos, et al., 2021; Di Francesco, Langtry, et al., 2023). (Acar and Solanki, 2009) applies **Vol** to a design optimisation problem in the automotive industry, quantifying the improvements in the reliability and weight of vehicle designs optimized for crash worthiness achieved by reducing uncertainty in structural material properties, crash performance estimates (from simulations), and experimental tolerances.

Few studies have used **Vol** to investigate how uncertainty reduction can improve decision making in energy systems. Almost all of these works look at capacity expansion planning problems in large-scale power generation or transmission-level energy systems, e.g. (Modiano, 1987), while this thesis focuses on planning in smaller, district-scale energy systems. For instance, (Krukanont and Tezuka, 2007) explores planning of the generation mix in the Japanese energy grid over a 10 year horizon. The key issue with all of these studies is that they analyse uncertainty reductions which are not achievable by any practical measurement. They all quantify the improvement in decision making achieved when the uncertainties in their problem are completely removed. This hypothetical case of obtaining perfect information provides an upper bound on the benefit of uncertainty reduction⁷. This can be useful for decision making if the upper bound is sufficiently tight, and in some sense quantifies the sensitivity of the decision making process to the uncertainties, allowing comparison and the most important uncertainties to be identified. Only one study (Wendling, 2019) finds one reasonably tight bound, and can conclude that the learning rate for the capital cost of CCUS-coal does not have a significant impact on planning the

⁷ This is discussed in more detail in Section 2.4.

decarbonisation of the global electricity sector. In fact, none of the uncertainties considered in these studies can be physically measured, for example the proportion of renewable generation in the future energy mix (Fürsch et al., 2014), the ultimate social cost of carbon (Wendling, 2019), or the future national transport energy demand (Krukanont and Tezuka, 2007). While research could improve the estimation of these quantities, the uncertainty reduction achieved will be far from total as these future values are inherently uncertain. As a result the studies can only provide rough suggestions as to at most how much money we should be willing to spend on research to improve the design of large-scale energy systems.

Only one existing study, (J. Niu et al., 2023), has applied Vol in the context of a district energy system. It looks at the sizing of chillers, gas generators, solar panels, and energy storage, in a district of five industrial buildings located in Shanghai to minimise the total cost (capital investment plus operation) of providing the cooling and electrical loads. Uncertainty in the solar generation and cooling & electrical loads are considered⁸. This work does not acknowledge the existing Vol framework and mature literature applying it in other fields. It does however compute the average reduction in the total cost of the energy system if all uncertainty is removed from the problem (which is called the **Expected Value of Perfect Information (EVPI)**, defined in Section 2.4). The key limitation of this study is that this uncertainty reduction is not related to any practical measurement that could be taken in the energy system. In fact, the uncertainty in solar generation as defined in the model used is not reducible, as at the time of system design the patterns of solar generation are not knowable. The uncertainty in the cooling and electrical loads also cannot be fully removed, as while improved information about the energy usage behaviours of the buildings could be gathered, their exact energy usage during operation also cannot be known. As with the large-scale system studies, the upper bound on the benefit of uncertainty reduction computed is large and so is not very informative.

So far no study has quantified the benefit that uncertainty reduction from practical data collection would provide to decision making in energy systems.

⁸ However, there are significant issues with either the methodology or the explanation of how the probabilities of these uncertainties are assigned and treated during the analysis.

1.2.4 Data collection requirements & costs

The importance of collecting data to support decision making in energy systems has been fully accepted for some time (Kathirgamanathan et al., 2021; Molina-Solana et al., 2017; Rysanek and Choudhary, 2013; Zhan and Chong, 2021). Recently there has been significant interest in the research community in greatly increasing monitoring of buildings to gather greater volumes and new types of data, and develop ‘smart buildings’ (Al Dakheel et al., 2020; Hernández et al., 2024). At the same time, many significant issues with data collection are widely acknowledged. Such as data quality, sensor reliability and the cost of maintenance, data relevance and curation, data warehousing and management, and data privacy (Mantha et al., 2015; Mobaraki et al., 2022; Xia et al., 2014). These issues reduce the ability of the collected data to support decision making, and overcoming them is costly. A prime example of the high cost of achieving building smartness is the UK smart meter rollout. Currently only 51% of electricity and gas meters in the UK are ‘smart’ (able to collect hourly energy usage data), despite the government’s rollout beginning in 2012, with a current expected total cost of £19.4bn (DESNZ, 2023f).

Few studies have looked at the data collection requirements for supporting decision making such as the design and control/operation of building energy systems.

(Zhai and Salazar, 2020) reviews case studies from the literature where submetered building energy usage data is used to identify energy savings measures. It investigates the relationship between the depth of submetering, the resolution of the building energy data used for analysis, and the energy savings achieved. It finds that submetering energy usage at higher spatial resolution enables greater energy savings down to the equipment level, but that submetering within plant equipment (sensor level) did not lead to additional energy savings. This indicates that building monitoring at the equipment level is required to properly identify energy savings measures. The average cost savings achieved are also quantified for each submetering resolution and for metering different equipment types. These costs are compared to the average hardware cost of the metering systems to assess the cost effectiveness of the different monitoring options. The cost savings from metering are found to be larger than the hardware costs, but this neglects a significant number of other costs associated with data collection, such as the costs of installing, maintaining, and using the monitoring systems. Another limitation is that only case studies are considered, and so the results are only valid for the types of building energy systems covered in those case studies. This work does not provide a generalisable method for assessing data collection requirements for new systems, new data variables, or even existing systems under altered conditions.

(Winschermann et al., 2023) investigates the impact of information availability on the performance of EV charging strategies in a set of simulated office buildings, with respect to satisfying user charging demands and limiting power draw from the grid. It quantifies the improvements in charging performance achieved when historic data on EV charging behaviours for the specific office building are available for planning, including the consideration of combinations of data variables, and the case where perfect foresight of EV behaviour is used. The analysis performed has some similarities to EVPI calculations, however it does not follow a clear decision making framework. The results demonstrate that using historic EV charging session data from the specific office buildings significantly improves the ability to set EV charging strategies in terms of both user service and peak power management, and so provides some indication that this could be worth collecting. However, the costs of acquiring and using this historic data to optimise charging are not quantified, nor is the charging performance of the energy system linked to a practical cost, e.g. the cost of grid connection or the lost revenue from unfulfilled charging demand. It is also shown that perfect foresight of EV charging requirements can further improve performance. However, this is not a practically available option, and so does not inform us of the actual benefit of using real-time monitoring to optimise charging strategies, which is very feasible to study. Therefore, it is difficult to use these results to guide decision making about which data is worthwhile collecting for developing EV charger controllers.

Far more work has been done to investigate data collection requirements in the contexts of model calibration, system identification, and forecasting. In the field of model calibration, studies have quantified the effects of the quantity of data gathered on the accuracy of calibrated energy models (Glasgo et al., 2017; Risch et al., 2021), the relative sensitivity of models to different types of energy data to prioritise collection (Tian, S. Yang, et al., 2016), data precision requirements and the cost of data collection (Chao Wang et al., 2022), and methods for maximizing the efficacy of collected data to reduce data collection costs (M. Han et al., 2021). Within systems identification, works have investigated the data requirements of different modelling approaches (Balali et al., 2023; Zhan and Chong, 2021; Q. Zhang et al., 2023), the impact of data resolution on model accuracy (Erfani, Jafarinejad, et al., 2023), the impact of model prediction accuracy on operational performance (Zhan, Lei, Jin, et al., 2022), and cost optimal data collection strategies to support model development (Zhan, Quintana, et al., 2022). And in forecasting, only one existing study, (Choi and S. Lee, 2023), has looked at the effect of the quantity of training data used on the accuracy of models forecasting thermal loads and building zone temperatures, which could be used for building control.

In these fields, the focus is on creating accurate models of energy systems. But ultimately these models are used to support decision making, and no studies have investigated how data collection to improve model accuracy carries through to improved decision making. Additionally, in very few of the studies listed above is the cost of collecting the data mentioned.

To determine if data is worth collecting, we need to understand how much it will benefit decision making (and so reduce the cost of what we're trying to do), and how much the data itself will cost to collect. However, discussions of data collection cost in the literature are often just holistic (Alahakoon and Xinghuo Yu, 2016). Other research has focused on reducing the quantity of data that needs to be collected in building energy systems, e.g. through the creation of standard datasets (Luo et al., 2022), but without discussing why it is costly or what those costs are. Where costs are quantified, the cost of hardware is often focused on (K. Han and J. Zhang, 2020), but this is just one of many important components of the total cost. For a building monitoring system there are significant costs from installing the sensors and data network needed, monitoring and maintenance to ensure sufficient data quality, and software licenses or consulting for managing, curating, and analysing the data. These costs are frequently neglected. For example, (Motegi et al., 2003) estimates the capital cost of electricity and gas smart-metering for a university campus to be \$0.27/m², plus an additional \$0.11/m² for maintenance and supporting IT systems. However, it does not account for installation costs, or cost of using the data to support building control, e.g. the substantial costs of training and deploying machine learning models for prediction (Strubell et al., 2020). (K. Han and J. Zhang, 2020) quotes the cost of sensors for occupancy detection to be \$0.2-3 per m² of building area, but acknowledges that many technologies suffer from either coverage or detection issues, which would need to be corrected at some expense.

1.3 Thesis overview

1.3.1 Research gap and thesis aims

The literature review in the previous section demonstrated the following research gaps regarding the use of data to support decision making in energy systems:

- 1 No studies have rigorously determined the data collection requirements for supporting the design and operation of energy systems, and how model accuracy affects the quality of decision making.
- 2 No studies have quantified the benefits of practical data collection for improving decision making by reducing uncertainty.
- 3 No studies have compared the benefits of practical data collection to its cost to determine whether it is worthwhile, and which data are most economical to collect.

The aims of this thesis are to:

- 1 Demonstrate that **Value of Information** analysis provides a principled and rigorous methodology for quantifying the benefit of data collection for supporting decision making in energy systems;
- 2 Investigate how training data impacts the accuracy of forecasting models, and how this affects the performance of **Model Predictive Control**; and
- 3 Study how uncertainty reduction from practical data collection affects the design of real-world energy systems, and so what data should be collected and how the decision making process should be adapted to make best use of that data.

These aims are achieved by:

- Chap. 2* Extending the **Value of Information** framework to allow the study of decision making
- Aim 1 via policies, and the benefit of retaining optionality as uncertainty reduces;
- Chap. 3* Illustrating how **Vol** can be used to quantify the benefit of data collection for supporting
- Aim 1 decision making across a range of example energy system decision problems;
- Chap. 4* Investigating the effects of training data on the accuracy of building load and grid
- Aim 2 condition forecasts, and the resulting performance of **MPC**;
- Chap. 5* Quantifying the value of hourly building load monitoring data for improving the design
- Aim 3 of solar-battery systems in a district energy system; and
- Chap. 6* Quantifying the value of building small-scale demonstrators, and retaining optionality
- Aim 3 in energy storage technology selection, for improving the design of an energy park.

1.3.2 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2 explains the [Value of Information](#) analysis framework, progressively building up the theory from first principles, and then extending it to study decision making via policies and the benefits of optionality. Alongside this, the interpretation of each analysis step and the insights the results provide on how uncertainties affect decision making in energy systems is illustrated by applying the theory to a simple worked example of designing a solar-battery system for a commercial building.

Chapter 3 then demonstrates how [Vol](#) can be used to understand the data collection requirements for supporting decision making in energy systems. It does so through three example decision problems, covering different building energy systems and decision making contexts. For each example, the model of decision making is explained, the [Vol](#) is quantified, and the insights this provides into the usefulness of collecting data are discussed.

Chapter 4 investigates how the available data affects the forecasting accuracy of machine learning prediction models of building load and grid conditions, and how the accuracy of these prediction models affects the performance of [MPC](#) in a simulated multi-building energy system. The data collection requirements for forecasting and measures to improve the data efficiency of prediction models are studied using a historic building energy usage dataset from the Cambridge University Estate. A case study approach is used rather [Vol](#), as the patterns in the training data are too complex for a Bayesian model.

Chapter 5 quantifies the value of using hourly load monitoring data to improve the sizing of solar-battery systems within a district, to determine whether uncertainty in building energy usage should be reduced before designing local generation & storage systems. It investigates a case study district energy system with grid constraints, modelled using the same building energy usage dataset from the Cambridge University Estate, where a Stochastic Program is used as a decision policy for system sizing.

Chapter 6 quantifies the value of retaining flexibility to adapt energy park designs and optionality over energy storage technology choice as uncertainty in storage performance reduces, to determine whether the energy park design process should be adjusted to enable this flexibility and optionality. An illustrative energy park model based on a proposed green hydrogen plant near Rotterdam is used, which considers four possible bulk energy storage technologies with uncertain round-trip efficiency, lifetime, and capital cost.

Chapter 7 concludes, summarising the results of the thesis, and discussing their limitations and how these could be addressed through future work.

1.3.3 Open data and models

All code and data used to perform the analyses in this thesis are publicly available under an open-source license. Links to the repositories are provided at the start of each chapter.

Additionally, as part of this PhD project, a dataset of energy usage data from buildings across the Cambridge University Estate was created and made publicly available. The dataset is described in Appendix A. It has been released as:

Max Langtry and Ruchi Choudhary (Apr. 2024). *Cambridge University Estates Building Energy Usage Archive*. doi: [10.5281/zenodo.10955332](https://doi.org/10.5281/zenodo.10955332). URL: <https://github.com/EECi/Cambridge-Estates-Building-Energy-Archive>

1.3.4 Research outputs and contributions

The work in this thesis has been published in the following articles:

Chapter 3

Max Langtry, Chaoqun Zhuang, Rebecca Ward, Nikolas Makasis, Monika J. Kreitmair, Zack Xuereb Conti, Domenic Di Francesco, and Ruchi Choudhary (2023). “Value of Information Analysis for Rationalising Information Gathering in Building Energy Analysis”. In: *Building Simulation 2023*. Vol. 18. Building Simulation. IBPSA, pp. 3483–3490. doi: [10.26868/25222708.2023.1478](https://doi.org/10.26868/25222708.2023.1478). URL: https://publications.ibpsa.org/conference/paper/?id=bs2023_1478

Max Langtry, Chaoqun Zhuang, Rebecca Ward, Nikolas Makasis, Monika J. Kreitmair, Zack Xuereb Conti, Domenic Di Francesco, and Ruchi Choudhary (Nov. 2024). “Rationalising Data Collection for Supporting Decision Making in Building Energy Systems Using Value of Information Analysis”. In: *Journal of Building Performance Simulation*, pp. 1–17. ISSN: 1940-1493. doi: [10.1080/19401493.2024.2423827](https://doi.org/10.1080/19401493.2024.2423827). URL: <https://doi.org/10.1080/19401493.2024.2423827>

Chapter 4

Max Langtry, Vijja Wichitwechkarn, Rebecca Ward, Chaoqun Zhuang, Monika J. Kreitmair, Nikolas Makasis, Zack Xuereb Conti, and Ruchi Choudhary (Oct. 2024). “Impact of Data for Forecasting on Performance of Model Predictive Control in Buildings with Smart Energy Storage”. In: *Energy and Buildings* 320, p. 114605. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2024.114605](https://doi.org/10.1016/j.enbuild.2024.114605). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824007217>

Chapter 5

Max Langtry and Ruchi Choudhary (Dec. 2025a). “Quantifying the Benefit of Load Uncertainty Reduction for the Design of District Energy Systems under Grid Constraints Using the Value of Information”. In: *Applied Energy* 400, p. 126549. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2025.126549](https://doi.org/10.1016/j.apenergy.2025.126549). URL: <https://www.sciencedirect.com/science/article/pii/S0306261925012796>

Chapter 6

Max Langtry and Ruchi Choudhary (Oct. 2025b). “The Value of Hedging against Energy Storage Uncertainties When Designing Energy Parks”. In: *Energy* 334, p. 137600. ISSN: 0360-5442. doi: [10.1016/j.energy.2025.137600](https://doi.org/10.1016/j.energy.2025.137600). URL: <https://www.sciencedirect.com/science/article/pii/S0360544225032426>

However, there is much more to a PhD than just a thesis. Alongside the work in this thesis, contributions⁹ were made to the following publications and projects:

Domenic Di Francesco, **Max Langtry**, Andrew B. Duncan, and Chris Dent (Sept. 2023). *System Effects in Identifying Risk-Optimal Data Requirements for Digital Twins of Structures*. doi: [10.48550/arXiv.2309.07695](https://arxiv.org/abs/2309.07695). arXiv: [2309.07695 \[stat\]](https://arxiv.org/abs/2309.07695). URL: <http://arxiv.org/abs/2309.07695>

Zoltan Nagy, Kingsley Nweye, Sharada Mohanty, Ruchi Choudhary, **Max Langtry**, Gregor Henze, Jan Drgona, Sourav Dey, Alfonso Capozzoli, and Mohamed Ouf (Dec. 2023). *The CityLearn Challenge 2023, NeurIPS*. URL: <https://neurips.cc/virtual/2023/competition/66590>

⁹ Starred items indicate a supervisory role in the project.

Wonjun Choi, Sangwon Lee, **Max Langtry**, and Ruchi Choudhary (Feb. 2025b). “Representation Vector-Based Time Series Similarity Analysis Using Unsupervised Contrastive Learning Time Series Encoder”. In: *Korean Journal of Air-Conditioning and Refrigeration Engineering* 37.2, pp. 72–81. issn: 1229-6422. doi: [10.6110/KJACR.2025.37.2.72](https://doi.org/10.6110/KJACR.2025.37.2.72). URL: <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE12036797>

* Jie Lu, Chaobo Zhang, Bozheng Li, Yang Zhao, Ruchi Choudhary, and **Max Langtry** (Feb. 2025). “Self-Attention Variational Autoencoder-Based Method for Incomplete Model Parameter Imputation of Digital Twin Building Energy Systems”. In: *Energy and Buildings* 328, p. 115162. issn: 0378-7788. doi: [10.1016/j.enbuild.2024.115162](https://doi.org/10.1016/j.enbuild.2024.115162). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824012787>

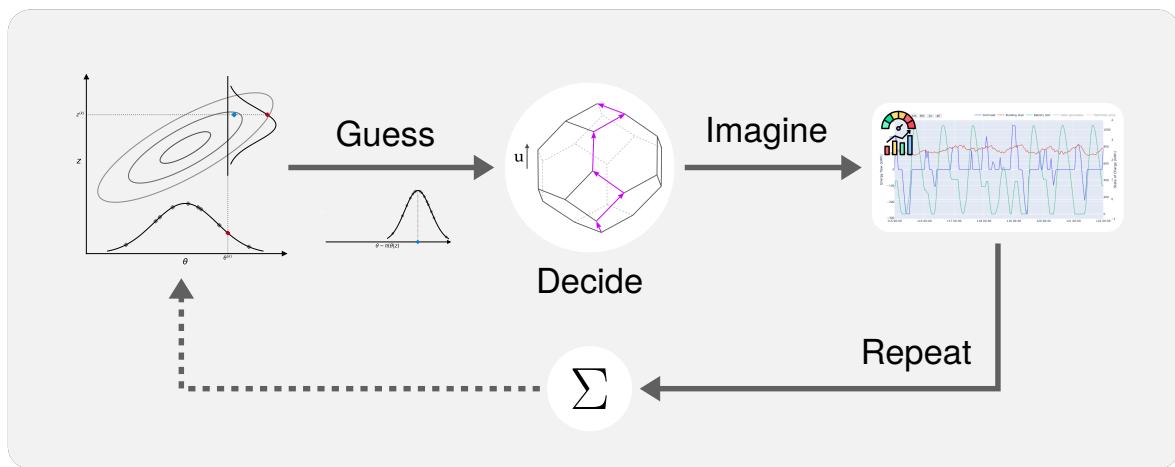
* Jie Lu, Zeyu Zheng, **Max Langtry**, Monty Jackson, Yang Zhao, Chenxin Feng, Ruqian Zhang, Chaobo Zhang, Jian Zhang, and Ruchi Choudhary (July 2025). *Automated Building Energy Modeling for Energy Retrofits Using a Large Language Model-Based Multi-Agent Framework*. (Submitted to iScience)

* Fabian Raisch, **Max Langtry**, Felix Koch, Ruchi Choudhary, Christoph Goebel, and Benjamin Tischler (Aug. 2025). *Adapting to Change: A Comparison of Continual and Transfer Learning for Modeling Building Thermal Dynamics under Concept Drifts*. (Submitted to Energy and Buildings). doi: [10.48550/arXiv.2508.21615](https://doi.org/10.48550/arXiv.2508.21615). arXiv: [2508.21615 \[eess\]](https://arxiv.org/abs/2508.21615). URL: <http://arxiv.org/abs/2508.21615>

Wonjun Choi, Sangwon Lee, **Max Langtry**, and Ruchi Choudhary (Sept. 2025a). *Privacy-Preserving Transfer Learning Framework for Building Energy Forecasting with Fully Anonymized Data*. (Working paper)

2

Value of Information analysis



Value of Information analysis (VoI) is the main methodology used in this thesis to investigate the benefit of data collection for improving decision making. This chapter explains the Value of Information framework. It begins with a brief introduction to VoI and its history in Section 2.1. The theory of the framework is then progressively built up through Sections 2.2 to 2.5. Alongside this theory, the interpretation of what the calculations are doing and their meaning is discussed, and a worked example of a simple building energy system is developed to illustrate how the results can inform decision making. Finally, the limitations of the framework are explored in Section 2.6.

2.1 Introduction

Value of Information analysis is a numerical framework for quantifying the expected benefit of collecting data to improve decision making by reducing uncertainty. The economic form was originally proposed by (Raiffa and Schlaifer, 1961)¹⁰, and is built on Bayesian Decision Analysis and Expected Utility Theory (Smith, 1945). It investigates how reducing epistemic uncertainty in a system (uncertainty associated with a lack of knowledge of a system (W.-H. Zhang et al., 2021)) by taking measurements affects the average utility/performance of the decisions within that system. It numerically answers the question,

How much better on average would a decision maker do at a given problem by taking a measurement to reduce uncertainty before making their decision?

This question has been and remains extremely important in lots of disciplines, and **Vol** has been used to study it in a wide variety of contexts. Early applications in the 1970s included (Conrad, 1980), operations management (Bedford and Onsi, 1966), agriculture (Perrin, 1976), accountancy (Mock, 1971), wargaming (Oldham, 1996), and hydrology (Klemeš, 1977). Due to the limited computational power available at the time, these studies used simple decision tree models of uncertainties and decision making. However, due to the interest in linear **Stochastic Programming** in Operations Research and Management Science at the time (Dantzig, 1955; Williams, 1965; Wilson, 1966) following the success of **Linear Programming** in supply chain management (Dantzig, 1956), a few studies applied **Vol** to decision making using small, early Linear Programs (Avriel and Williams, 1970).

More recent uses include fields such as environmental science, medicine, agriculture, and economics (Keisler et al., 2014). For example in medicine, **Vol** has been used to study the trade-off between reducing uncertainty in the efficacy of treatments and the cost of performing clinical trials (Willan and Pinto, 2005), and to decide whether further research is required before a new medical technology is adopted (Tuffaha et al., 2014).

Within engineering, **Vol** has been used in contexts such as maintenance scheduling for buildings (Grussing, 2018) and wind farms (Myklebust et al., 2020), construction project planning (Esnaashary Esfahani et al., 2020), automotive design (Acar and Solanki, 2009), sensor placement (Malings and Pozzi, 2016), and structural health monitoring (Di Francesco, Chryssanthopoulos, et al., 2021; Di Francesco, Langtry, et al., 2023).

Review papers (Keisler et al., 2014) and (W.-H. Zhang et al., 2021) provide thorough overviews of the applications of **Vol** across disciplines and in engineering respectively.

¹⁰ Though several other authors were working on similar ideas in other fields around the same time (Feltham, 1968; Howard, 1966; Hurley, 1964; Spivey, 1968; Szaniawski, 1967)

2.2 Statistical decision theory

To understand how uncertainty affects decision making in a quantifiable way, we first need a mathematical framework for making decisions. For this we'll use the mathematics of optimisation¹¹. Numerical optimisation provides us with a well-defined, objective, and consistent way to make decisions, which does not rely on unquantifiable human judgement¹².

Let's imagine we have some problem we want to solve¹³, and to do so we have to select one choice, which we'll call an 'action' a , from a set of possible actions, \mathcal{A} . How well each action does at our problem is described by a utility function, $u(a)$. Our actions can have multiple aspects (or dimensions), and the utility function returns a single number for each action which describes how 'good' it is at solving the problem. Costs can be described with negative utilities.

We want to find the best action to take, that is the one which achieves the highest utility. Mathematically we describe this optimisation problem¹⁴ as,

$$\max_{a \in \mathcal{A}} u(a) \quad (2.1)$$

How to actually search through the set of possible actions to find the optimal choice for a practical problem is a complicated question. There are many *many* different optimisation algorithms, all of which are suited to different situations (Bonnans et al., 2006). But for our purposes, let's assume we can choose a suitable method for our problem, and that we are able to properly optimise our utility to find the best action, which is the one we'll take.

This is a simple deterministic model of decision making which assumes we know exactly how the action we take will perform. However in practical engineering systems, there are usually factors which we don't precisely know at the time the decision is made that affect how the different actions will perform when taken, i.e. the utility they achieve. Because of these uncertain factors, which we'll refer to as the 'uncertainties' of the system, θ , there are many possible outcomes if a given action is taken. We can graphically represent this process of taking an action, realising the value of the uncertainties, and observing an outcome/utility, using a decision tree which is illustrated in Figure 2.1.

¹¹ In the energy systems field this is referred to as optimisation based design/control (Barber and Krarti, 2022)

¹² Of course the setup of our optimisation *is* determined by subjective human judgement. We need to define what we care about, i.e. what objective we want to optimise and what our constraints are. We can never really achieve complete objectivity. But, given a particular problem we want to solve, with a goal we can quantify, optimisation give us an objective tool for tackling it.

¹³ We are the decision maker, also called the 'actor'.

¹⁴ Mathematics uses somewhat different notation from decision theory to describe optimisation problems, see (Boyd and Vandenberghe, 2004).

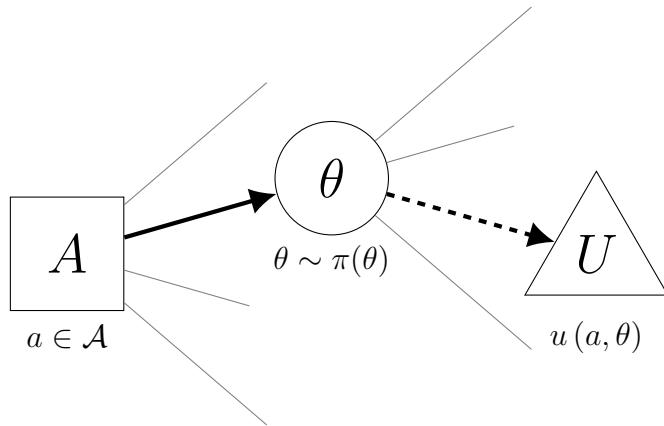


Fig. 2.1 Decision tree representation of decision making under uncertainty.

Adapted from Figure 1 of (Di Francesco, Chryssanthopoulos, et al., 2021)

While we don't know the value of the uncertainties, and so the outcome that we'll get, we typically do have probabilistic knowledge of them. This knowledge comes in the form of their probabilistic model or distribution^{15,16}, $\pi(\theta)$.

As the utility now depends on both the action and the value of the uncertainties, $u(a, \theta)$, there is now also uncertainty or risk associated with the utility we'll get, which we must manage during decision making. There are many ways of handling this risk, such as taking the action with the 'best worst-case' utility, or the one with the best robustness guarantees. But the most common way is to take the action with the highest average (or expected) utility¹⁷. So the 'objective' we now want to optimise to determine our action choice is the expected value of the utility over the uncertainties in the system,

$$\max_{a \in \mathcal{A}} \mathbb{E}_\theta \{u(a, \theta)\} \quad (2.2)$$

¹⁵ In this chapter will assume that you are somewhat familiar with probabilities, likelihoods, sampling from distributions, and calculating expectations. If not, there are many textbooks and courses that introduce these statistical ideas, such as (Schwarzlander, 2011).

¹⁶ The distributions of the uncertainties can either be modelled using existing data from similar system, or by making reasonable assumptions based off knowledge of the system. The choice of these distributions is a critical modelling step (van de Schoot et al., 2021).

¹⁷ The framework can be easily adapted to use whichever uncertainty management method is most appropriate for the decision making context. This is done by simply changing the objective. For example in Chapter 6, a **Conditional Value-at-Risk** term is added to the objective to incorporate risk aversion into the decision making. In fact, many risk management methods just adjust the utility function, adding a risk term, and still optimise the expectation.

However, for real engineering problems we typically cannot calculate this expectation exactly¹⁸. Instead, we use a Monte Carlo (MC) estimate to approximate it¹⁹. To do so, we draw a number of samples from the distributions of the uncertainties, $\theta^{(n)} \sim \pi(\theta)$, and compute the average utility over those samples,

$$\frac{1}{N} \sum_n u(a, \theta^{(n)}) \quad (2.3)$$

As the number of samples used, N , increases, this approximation improves, and the estimate converges to the true value of the expectation.

So to solve our decision making problem under uncertainty, we optimise the average utility achieved over the sampled values of the uncertainties,

$$\max_{a \in \mathcal{A}} \mathbb{E}_\theta\{u(a, \theta)\} \approx \max_{a \in \mathcal{A}} \frac{1}{N} \sum_n u(a, \theta^{(n)}) \quad (2.4)$$

Doing so tells us the best action to take, a^* , and how ‘good’ we expect it to be at solving our problem, i.e. the average utility it achieves. We can also use the distribution of its utility over the uncertainty samples to understand the risk associated with our decision making.

More generally, this type of optimisation is referred to as a ‘stochastic decision problem’. It is fully specified by the set of possible actions, \mathcal{A} , the probabilistic model of uncertainties, $\pi(\theta)$, and the utility function²⁰, $u(a, \theta)$. Solving the stochastic optimisation in Eq. 2.2 gives us the best solution to this problem. But solving it can be very computationally expensive, as the utility function needs to be evaluated N times due to the sampling²¹. Instead, we could try to solve the problem using some heuristic (Pickering, Ikeda, et al., 2016). A common approach is to use the expected (average) values of the uncertainties in a deterministic optimisation, which is called the Expected Value Problem (EVP). This is a much cheaper problem to solve, but may not correctly identify the best action to take. The difference in expected utilities achieved by the stochastic optimisation (Eq. 2.2) and the solution to the EVP tells us how much better we can do at solving the decision problem by accounting for the uncertainties in the system in our optimisation, and is referred to as the Value of Stochastic Solution (VSS) (Bistline, 2015). Actually, we could use whatever policy²² we like to try to solve the decision problem. This more general setting is discussed in Section 2.5.1.

¹⁸ The expectation can only be calculated exactly if the probabilistic model of the uncertainties used is discrete, or if the utility function is very simple.

¹⁹ In this thesis, whenever there is an expectation, we will use an MC estimate of it in our numerical calculations.

²⁰ Strictly, it’s the optimisation objective which is a function of the utility that defines the problem, but these are often used interchangeably.

²¹ In fact for a Stochastic Program, the computational cost of using this average utility scales as $\mathcal{O}(N^3)$.

²² The term ‘policy’ is used to refer to any structured way of making decisions. E.g. a flow chart defines a policy.

2.E.1 Introducing an example problem

To help clarify this discussion of mathematics and understand why it's useful to us, let's apply it to something concrete and introduce an example problem. We'll developed this example problem throughout the chapter, and demonstrate how we can use each step of the mathematical framework as we introduce it. Code for all the calculations in this example is available for you to try out and explore at <https://github.com/mal84emma/VoI-Tutorial>.

Designing a solar-battery system for a commercial building

Let's imagine we want to decarbonise the energy use of a hypothetical commercial building. To do this we plan to install solar panels and battery storage in the building. This equipment is expensive, but by generating our own energy and doing some energy trading with the grid we can reduce the electricity bill. So our goal is to design the solar-battery system for the building, that is choose the sizing of solar panels and battery storage, to minimise the total cost of meeting the building's energy demands.

To help make this design decision we'll use a simple model of the building energy system to calculate the overall cost of constructing and operating the system, which is described in Appendix 2.A.

Initially, we'll optimise the system design by looking at 5 design options selected using our experience of building energy systems, and run the energy model to identify the design with the lowest cost. These design options are chosen to generate between 66% and 100% of the building's annual energy demand with the solar panels, and to use between a 4hr and 8hr battery to capture and shift the solar peak. The cost of each design is determined by running the model simulation using data from a single design year, in this case 2017. We can compare these costs in Table 2.1.

The results show us that there is a trade-off between the capital and operating costs, and that the design with 750kW of solar panels and 600kWh of battery storage gives us the best balance. With this design, it costs us £152.3k/yr in total to supply the building load, equivalent to £0.174/kWh. If the solar-battery system wasn't installed, buying all the electricity from the grid would cost £219.2k/yr (or £0.250/kWh), so we save £66.9k/yr (a 31% cost reduction) by installing the system, as well as significantly reducing our carbon emissions.

Solar capacity (kWp)	Battery capacity (kWh)	Capital cost (£k/yr)	Operating cost (£k/yr)	Annualised cost (£k/yr)
<i>Discrete design options</i>				(Discrete optimisation)
500	400	65.5	100.6	166.1
500	600	79.5	91.4	170.9
600	600	87.0	76.1	163.1
750	600	98.2	54.1	152.3
750	800	112.2	46.3	158.6
<i>Linear Program solution</i>				(Continuous optimisation)
889	501	101.7	39.3	141.0

Table 2.1 Deterministic design optimisation results.

We have in some sense optimised this system design, but we've only looked through a very small number of options. We could search through lots more options, but this would be very computationally expensive. However, our building energy system model actually uses a Linear Program ([LP](#)) to optimise the operation of the system and find lowest cost of running it. So, we can include the solar and battery capacities as decision variables in our [LP](#) optimisation, and use the nice mathematics of our model to find the best capacities out of all possible options, and so the 'optimal' design. This solution is also shown in Table 2.1. Using the [LP](#) leads us to a design we might not have chosen to explore. It generates more solar energy than the building can use, and is more cautious with the batteries than we were. And importantly, it reduces the cost by 7.4% compared to our best design guess.

This design optimisation is all well and good, but we've completely neglected the fact that there are many factors that will affect the operation, and so cost of running of our energy system, that we don't know at the time we design it. So let's introduce some uncertainties into our model, and see how they affect our design decisions. We'll account for uncertainties in the building load, solar generation, and the cost and efficiency of the battery storage. The probabilistic models we'll use are described in Appendix 2.A.2.

We can repeat the optimisations from before, but now we will draw 50 samples^a from the distributions of the uncertainties, and our design objective will be the average cost from the energy model over these samples. We call a set of values sampled from the uncertainties a 'scenario'.

Solar capacity (kWp)	Battery capacity (kWh)	Av. capital cost (£k/yr)	Av. operating cost (£k/yr)	Av. annualised cost (£k/yr)
<i>Discrete design options</i>				(Discrete optimisation)
500	400	65.5	95.7	161.2
500	600	79.6	86.4	166.0
600	600	87.1	71.1	158.1
750	600	98.3	49.1	147.4
750	800	112.3	41.2	153.6
<i>Stochastic Program solution</i>				(Continuous optimisation)
807	429	90.5	49.4	139.9

Table 2.2 Stochastic design optimisation results.

Comparing to the results from Table 2.1 for the discrete optimisation, we can see that the same design option is selected as before, but that there is about a 3% difference in the estimated cost of the system. I.e. using the single design year causes us to overestimate the cost of the system compared to the actual average.

Designing using the LP, which is now a Stochastic Program^b (SP), leads to a design which uses around 10% smaller capacities than the deterministic case, but doesn't cause a significant change in total cost between the designs^c.

So in this case, accounting for these pretty significant uncertainties in our optimisation model doesn't have a substantial impact on the design decision^d, but it does provide us with a better estimate of how much the system might cost^e.

Another advantage of accounting for uncertainties in our design model is that we also get information about the resulting uncertainty in the cost, which can be useful for planning and managing risk. Over the 50 scenarios, the total cost of the 750kWp–600kWh system varied from £110.4k/yr to £175.5k/yr, with a standard deviation of £16.6k/yr. Knowing that the cost could get this high would be important if we had a strict budget constraint, and might cause us to choose a different design^f. Or ask for a bigger budget.

^a This isn't that many samples, and so our estimate of the expected cost might have some noisy/inaccurate, but this is the largest number we can use within an achievable computational cost.

^b Specifically, it's a linear scenario program, where the (linear) operating constraints of each scenario are all included in the model, and the objective is the average total cost over the scenarios. See Eq. 2.20.

^c Unfortunately we cannot actually compute the VSS as the deterministic LP solution is infeasible, which is a problem, but we can approximate it. This issue is discussed further in Appendix 2.A.3.

^d Though this might depend on the design year we chose. We may have just got lucky.

^e Not only is the average cost value more accurate, but the statistical information about the possible costs is also useful for planning.

^f If we were very risk averse, we might want to optimise for the maximum cost, rather than the average.

2.3 Bayesian decision analysis

While we are able to manage uncertainties during our decision making, their presence is problematic for decision making. This is because when we are hedging our decision to perform well on average, we're missing out on making the best decision for the state of the system that we'll actually get (the true values of the uncertainties). Though of course at the time we are making our decision we don't know what that will be. But, if we were able to reduce the uncertainty in our problem²³, and get better information about the true state of the system, we would need to hedge less, and so could do a better job of making our decision.

Let's consider how our decision making will change if we are able to take a measurement/collect data that reduces the uncertainties before we have to make our decision.

First of all we need to be able to describe how the data informs us of the true state of the system and reduces the uncertainties. We do this using a likelihood function. For a measurement, e , the likelihood function, $f_e(z|\theta)$, describes the chance of getting a particular value/data from the measurement, z , if the true value of the uncertainties is θ . This is to say it describes the uncertainty associated with the measurement²⁴.

Bayes' theorem tells us the following,

$$p(\theta | z) = \frac{p(z | \theta) p(\theta)}{p(z)} \propto p(z | \theta) p(\theta) \quad (2.5)$$

where $p(\theta)$ is the prior distribution of the uncertainties, $p(z | \theta)$ is the likelihood function of the measurement, and $p(\theta | z)$ is the posterior distribution of the uncertainties for a given measurement value/data z .

The prior distribution describes our initial knowledge of the uncertainties. The likelihood function describes the knowledge/information we gain from the data. And the posterior distribution combines these two pieces of information, and tells us about the remaining uncertainty given what we've observed.

Though the framework is typically described in terms of physical measurements, we can apply these ideas to any source of information that reduces the uncertainty in our problem, such as research, improved forecasting methods, or getting expert advice. The

²³ Not all uncertainties can be reduced. There are two types of uncertainty. Aleatoric uncertainty, which is inherent variability which cannot be explained away, and epistemic uncertainty, which is uncertainty associated with a lack of knowledge of the system, which can be reduced through data collection. Often in engineering systems uncertain parameters have a mix of aleatoric and epistemic, meaning their uncertainty can be reduced but never fully removed.

²⁴ A standard likelihood function is a Normal/Gaussian with a precision/variance that is determined from lab experiments, e.g. $\pm 10\%$ meaning the standard deviation would be 5% of the mean (for a 95% confidence level).

only requirement is that we can model the uncertainty reduction that information provides²⁵, i.e. come up with a likelihood function.

As the Vol framework is numerical, we can use a probabilistic programming language such as Stan (Carpenter et al., 2017) to sample from the required distributions, giving lots of flexibility in how we can model the uncertainties and how they are reduced by data collection.

At the time we're considering whether to take a measurement to support our decision making, we can't know what value/data, z , we'll get from the measurement. This means there are many possible outcomes for each available measurement option. Once we've received the data from our measurement, we can use it to form a posterior distribution of the uncertainties, $\pi(\theta|z)$, which has reduced uncertainty compared to the initial, prior distribution. We use this improved information to support and improve our decision making, performing the same stochastic decision process as before, but now using the updated knowledge of the possible outcomes of our actions. We can represent this extra measurement step in the decision making process by extending the decision tree, as shown in Figure 2.2.

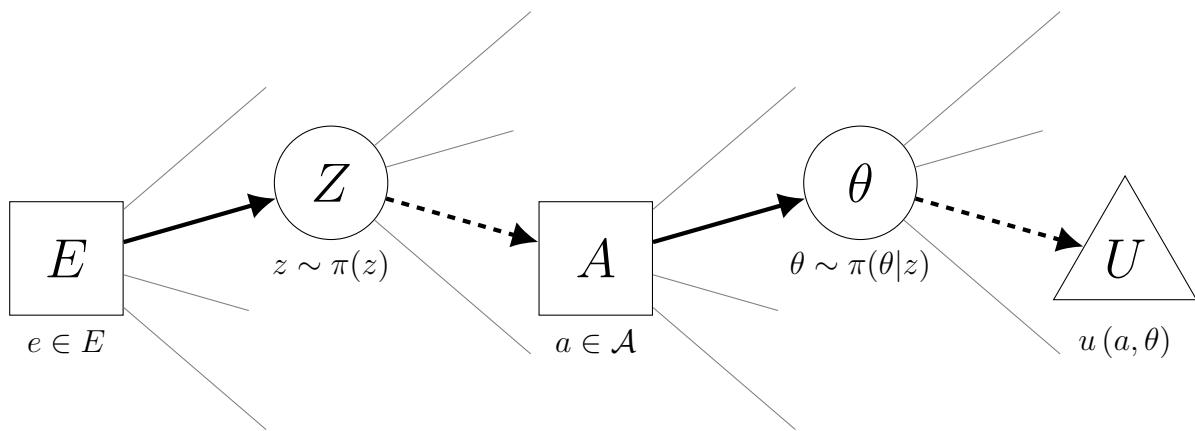


Fig. 2.2 Decision tree representation of decision making with measurements

Adapted from Figure 1 of (Di Francesco, Chryssanthopoulos, et al., 2021)

While these decision trees illustrate the sequence of our decision making, the available choices and their possible outcomes, they are only really useful for illustrating simple toy problems. For complex, real world decision making problems, we use influence diagrams to provide a clear and concise visual description of the structure of the system and decision making. These diagrams show only the causal dependencies between decisions, uncertainties, and outcomes in the system, and allow domain experts to quickly describe decision

²⁵ Though if modelling this uncertainty reduction is tricky, we can make a best guess and use sensitivity analysis to see if that guess affects our conclusions.

making in complex engineering systems. They are excellently explained by (Di Francesco, 2023).

To understand how collecting this data affects our decision making, we can update our decision optimisation to use this improved information that reduces the uncertainties in the problem,

$$\max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \quad (2.6)$$

This new optimisation (stochastic decision problem) can be used to find the decision we would make if we'd received data z . We refer to this as the Posterior decision problem, as we are now finding the action that performs best on average under the remaining uncertainty given by the posterior distribution, $\pi(\theta|z)$ ²⁶.

However, because we don't know what data z we'll receive before we measure, to understand how well we would do using data to support our decision making, we need to average over the possible values of the measurement data²⁷,

$$\mathbb{E}_{\theta,z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} \quad (2.7)$$

This tells us how well we would do on average at solving our decision problem, i.e. the expected utility we'd achieve, if we were to gather data to reduce the uncertainties in the problem before making our decision.

We refer to this as the Pre-Posterior decision problem, as we are averaging over all of the possible Posterior decision problems corresponding to each measurement value z . Eq. 2.2 from before is referred to as the Prior decision problem, as there our decisions are made with only the information from the prior distribution, $\pi(\theta)$.

When solving the Pre-Posterior decision problem numerically, we first hypothesis possible values z we could receive from the measurement by sampling from the prior and likelihood functions. Then, for each sampled z value, we consider what the corresponding Posterior decision problem with reduced uncertainty would be, how we would solve it, and how well we would do. Finally, we average those expected utilities to estimate the overall expected utility over solving the decision problem with the help of data collection.

²⁶ Note, in Eq. 2.6 we use ϑ to denote estimates of the true value of the uncertainties which are made after data z is received, to differentiate from the possible true values of the uncertainties θ that we will later average over.

²⁷ We also explicitly average over the possible values of the uncertainties, θ , as the measurement values are associated with the true states of the system. During sampling, we first sample θ values from the prior, and then sample the corresponding likelihood function to obtain z values.

$$\frac{1}{N} \sum_{z^{(n)} \sim f_e(z|\theta^{(n)})} \max_{a \in \mathcal{A}} \left(\frac{1}{M} \sum_{\vartheta^{(m)} \sim \pi(\vartheta|z^{(n)})} u(a, \vartheta^{(m)}) \right) \quad (2.8)$$

This means we now get a distribution of the actions we would take given the possible measured values, and how well we would do at solving our problem in each case.

Collecting data to reduce the uncertainty in the problem will improve our decision making ability. This is because the decisions we now make are more tailored to the actual state of the system (which we now know more precisely) and use less hedging. We could always take the decision we would initially have made (i.e. our choice from the prior decision problem), but if we see certain data values we can adjust our decision to take a choice better suited to the likely state of the system. We can demonstrate this improvement from uncertainty reduction mathematically,

$$\mathbb{E}_{\theta, z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} > \max_{a \in \mathcal{A}} \mathbb{E}_{\theta, z} \{ \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \} = \max_{a \in \mathcal{A}} \mathbb{E}_{\theta} \{u(a, \theta)\} \quad (2.9)$$

In the more general case where we have a set of measurement options, E , are available to take, we can also optimise our choice of measurement,

$$\max_{e \in E} \mathbb{E}_{\theta, z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} \quad (2.10)$$

2.E.2 Measuring building energy usage

Previously in our example problem, we acknowledged that when we design the solar-battery system, we don't know exactly the amount of energy the building will use during operation. And we accounted for this uncertainty in the building load in our design decision, as well as other uncertainties, finding the design that provides the lowest cost on average.

Our hypothetical commercial building has a design load of 100kW, and from our experience of previous buildings, we expect that the load during operation will be within $\pm 20\%$ of this value. So we modelled this uncertain load using a Normal distribution

truncated at 2 standard deviations,

$$\theta_{\text{load}} \sim TN(\mu = 100, \sigma = 10, a = 80, b = 120)$$

This distribution of the load uncertainty, the prior distribution, describes our initial knowledge of the building load, also called our prior knowledge/belief.

We can reduce this load uncertainty by gathering data from the building. For example by collecting monthly energy usage data from the building's electricity meter. Again from our experience of other buildings, we know that the energy usage of a building in future years is likely to be within $\pm 5\%$ of the previous year's usage^a. So we can describe the information we'd get from this metering data using the following likelihood function,

$$z_{\text{load}} | \theta_{\text{load}} \sim \mathcal{N}(\mu = \theta_{\text{load}}, \sigma = \varepsilon \theta_{\text{load}})$$

where θ_{load} is the uncertain operational building load, z_{load} is the previous year's load we get from the metering data and use to estimate future loads, and $\varepsilon = 0.025$ ^b.

The uncertainty reduction that collecting this metering data provides is illustrated in Fig. 2.3, which shows the prior distribution of the building load during operation, an example annual load value (measurement) we might get from the electricity meter, and the corresponding posterior distribution.

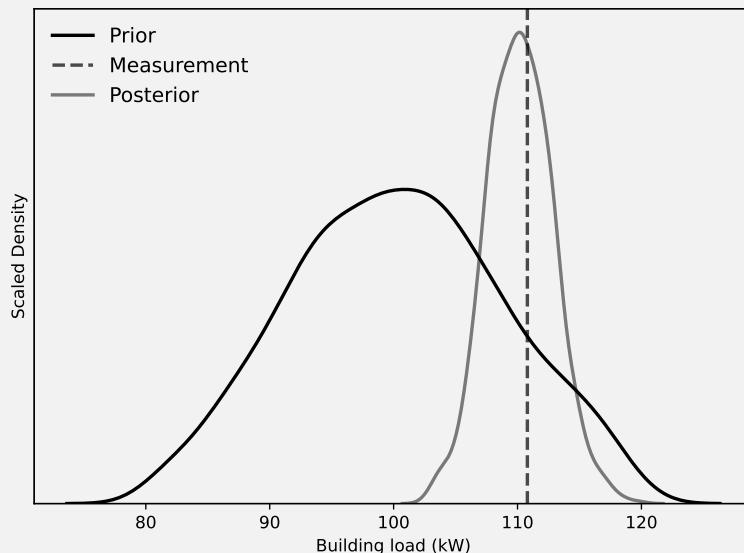


Fig. 2.3 Example reduction in load uncertainty provided by metering data.

Reducing uncertainty in the building load gives us a better understanding of whether energy we generate with the solar panels can be used by the building, and how much energy we need to store in the batteries to run the building during peak electricity price periods. So, it reduces uncertainty in the operational cost of system designs. This allows us to make a design decision that is better suited to how the building will actually behave.

We can represent the structure of the solar-battery system design process and how gathering monitoring data affects our decision making using an influence diagram.

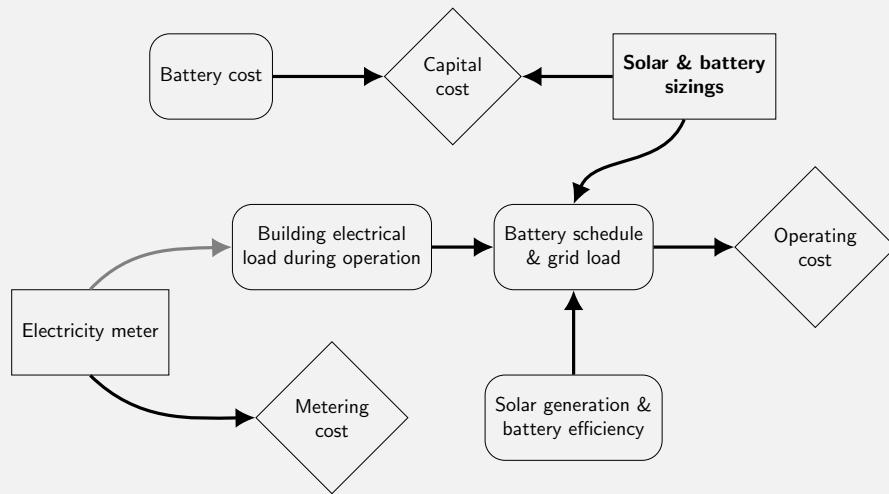


Fig. 2.4 Influence diagram representation of example energy system design problem.

Square nodes represent decisions, rounded nodes represent uncertainties or uncertain factors, and diamond nodes represent outcomes or utilities/costs. Arrows represent causal dependencies between nodes, with grey arrows indicating reduction in uncertainty from data collection.

To understand how collecting the metering data and reducing load uncertainty might affect our design decision, we can look at the distribution of designs we would choose using our **SP** given the possible measurement values we might get. To do so, we hypothesise 50 possible true building load values, θ_{load} , by drawing from the prior load distribution. We then hypothesise a possible measurement (previous annual load values we get from the metering data, z_{load}) for each true load value by drawing a sample from the corresponding likelihood function^c. The information from each measurement is combined with the prior to form a posterior. And for each measurement value, we perform the design optimisation process by sampling load values from the corresponding posterior, and values for all other uncertainties from their priors as we assume we're not getting information about them, then using the **SP** to optimise over these 'posterior scenarios' which have reduced load uncertainty.

Fig. 2.5 shows this distribution of ‘posterior designs’. Across the designs there is little change in the solar capacity, around $\pm 5\%$, however the battery capacity chosen varies by $\pm 25\%$. Before, we saw that accounting for uncertainty during the design process didn’t have much affect on the chosen design. That is to say that the design that performs best *on average* was similar to the best design for the average scenario^d. However we can clearly see that the uncertainty in building load *does* have a significant effect on the best design for the actual building.

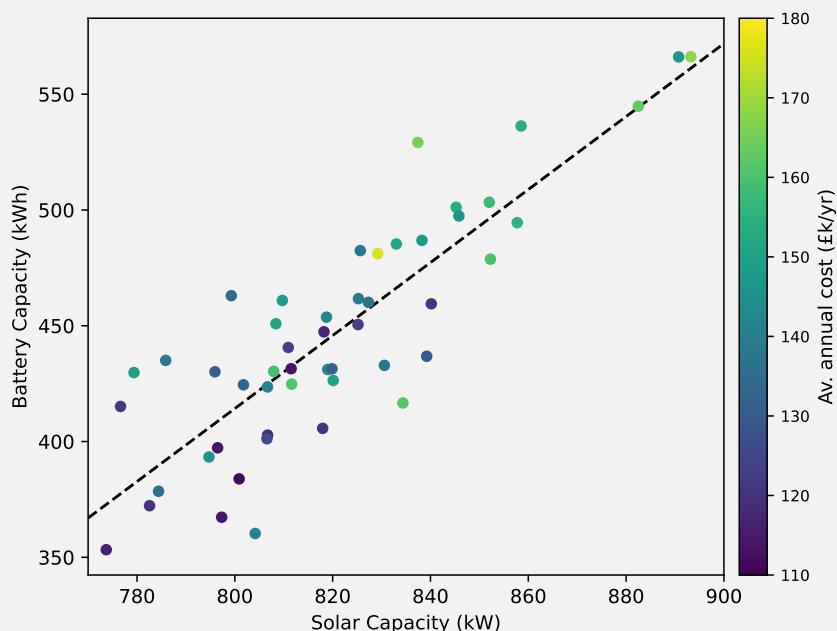


Fig. 2.5 Distribution of solar-battery system designs made with metering data.

^a Assuming there are no significant changes to the building, e.g. changes of usage or retrofits. At the time of designing the solar-battery system we do not expect any such changes, as heat pumps have recently been installed to decarbonise the building’s heating.

^b To put our experience of future years’ energy usage being within $\pm 5\%$ of the previous year’s within a 95% confidence interval.

^c This is the same as drawing from the joint prior distribution of θ and z .

^d Specifically, we used the average annual load value in our deterministic optimisation. We’re not sure how ‘average’ the year of solar and load profile data we used is. Though there methods for coming up with an ‘average year’ of data, and these are commonly used for energy system design.

2.4 The Value of Information

The **Value of Information (Vol)** is defined as the difference in the performance of decisions made with and without extra information/uncertainty reduction. So, it is given by the difference between the expected utilities achieved by solving the Prior decision problem (Eq. 2.2) and the Pre-Posterior decision problem (Eq. 2.7). It numerically answers the question,

How much better could we do on average if we collected data to reduce the uncertainties in our problem before making a decision?

If the decision problem has an economic objective (i.e. the utility is a cost or a profit), the **Vol** tells us how much we should be willing to pay for the data that reduces the uncertainties in our problem. By comparing the **Vol** to the cost of data collection, **we can determine if collecting that data is actually worthwhile for supporting our decision making.**

2.4.1 Perfect Information

Initially let's consider the simplest case of **Vol**, where we assume we have the ability to perfectly measure the system and remove all uncertainty from our problem. This means the data we collect tells us the true values of the uncertainties, $z = \theta$. So in this case the Pre-Posterior decision problem can be simplified, as the Posterior decision problem is now deterministic, we assume we know the true state of the system we want to optimise for,

$$\mathbb{E}_{\theta,z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{ u(a, \vartheta) \} \right\} \rightarrow \mathbb{E}_{\theta} \left\{ \max_{a \in \mathcal{A}} u(a, \theta) \right\} \quad (2.11)$$

We find the **Vol** by comparing the average utility achieved by these decisions made with perfect information to the average utility of the Prior decision problem, where the decision is made without any extra information about the uncertainties. In this case the **Vol** is referred to as the **Expected Value of Perfect Information (EVPI)**,

$$\text{EVPI} = \mathbb{E}_{\theta} \left\{ \max_{a \in \mathcal{A}} u(a, \theta) \right\} - \max_{a \in \mathcal{A}} \mathbb{E}_{\theta} \{ u(a, \theta) \} \quad (2.12)$$

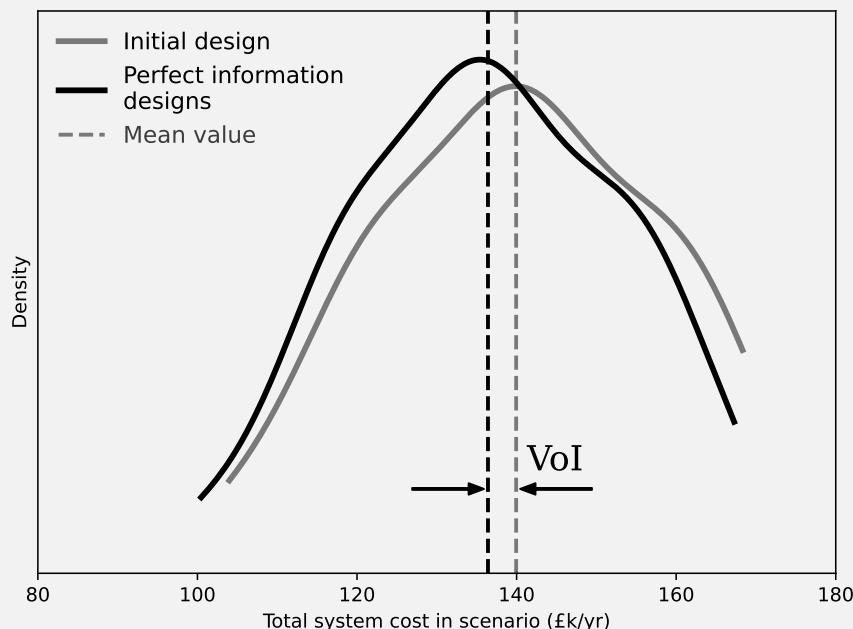
The **EVPI** provides an upper bound on all the other **Vol** values. It is also much cheaper to compute as the posterior optimisations are deterministic. However **EVPI** values can be misleading, as it's not practically possible to completely remove uncertainty, and so we cannot know how much of that value can actually be obtained.

2.E.3 Value of removing uncertainty

Let's imagine what would happen if we were able to perfectly know everything about our building and the solar-battery system we'll put into it, to see how much better we would do if we could achieve the theoretically best possible design. Of course this is unrealistic. Some of the uncertainties in the system couldn't be reduced, such as the pattern of solar generation during operation, and for the others, the uncertainty could never be completely removed. For example, we can't know the exact amount of energy the building will use during operation when we design it. But this thought experiment will give us some insight into how much these uncertainties are affecting our ability to make our design decision.

To do this, we perform a similar process to [2.E.2](#), but we now sample scenarios from the prior distributions of all the uncertainties, and then use the [SP](#) to find the best design for each scenario, assuming that in each hypothesised case we knew the true values of the uncertainties.

[Fig. 2.6](#) compares the distribution of costs achieved by the initial/prior design (from [Table 2.2](#)) over the scenarios, to the costs of the best possible design for each scenario (the designs made with perfect information).



[Fig. 2.6](#) Comparing distributions of possible system costs for designs made with and without perfect information.

The average cost of the designs made with perfect information is £136.4k/yr. Comparing to the £139.9k/yr average cost of the initial design from Table 2.2, the **EVPI** is £3.5k/yr, or 2.5% of the prior cost.

In the context of improving energy system designs to reduce energy costs this is a pretty significant saving. However the **EVPI** is only an upper bound. So we can't know how much of this cost saving we could actually achieve by collecting data in practice. All we can conclude from this is that reducing the uncertainties in our problem could save us *up to* £3.5k/yr. So it *might* be worth doing something about the uncertainties to improve our design decision. Or it might not.

2.4.2 Partial Perfect Information

Often not all of the uncertainties in a system are, or even can be, reduced (see footnote 23). So a more realistic case is one where some uncertainties are removed, but some others remain. This means in the Posterior decision problems, as not all uncertainties are removed, we need to optimise the average utility over the remaining uncertainties. Let's call the uncertainties that are removed by the data collection θ , and the ones that remain unchanged ϕ . In this case the **Vol** is referred to as the **Expected Value of Partial Perfect Information (EVSSI)**, and is given by,

$$\text{EVSSI} = \mathbb{E}_\theta \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_\phi \{ u(a, \theta, \phi) \} \right\} - \max_{a \in \mathcal{A}} \mathbb{E}_{\theta, \phi} \{ u(a, \theta, \phi) \} \quad (2.13)$$

2.E.4 Designing with known load

Now, let's assume that we only removed the uncertainty in the annual energy usage of the building (or mean load), meaning that we knew perfectly at the time of design the aggregate energy demand we'd need to meet. In this case, the average cost of our designs made with perfect mean load information would be £139.4k/yr. So the **EVSSI** is just £0.5k/yr, or 0.4% of the prior cost.

This is a much smaller value than the **EVPI** we calculated before, which tells us that the other uncertainties we're no longer removing^a have quite a big impact on our design decision. So removing this pretty significant uncertainty ($\pm 20\%$) in the mean building load doesn't provide us that much of a cost saving for our design. We may decide that it's not worth chasing this saving, and so the upper bound we get from this simplified case can tell us that this information isn't worthwhile collecting.

^a Which are in the cost and efficiency of the batteries, and the amount and pattern of solar generation.

2.4.3 Imperfect Information

In the general case, collecting data reduces but does not remove uncertainties²⁸. We refer to this as gathering ‘imperfect information’. How we model uncertainty reduction from data and understand its effect on decision making is discussed in Section 2.3. If we have some measurement option e available, which has likelihood function $f_e(z|\theta)$, the value of the information provided by that measurement is referred to as the **Expected Value of Imperfect Information**²⁹ (**EVII**), and is given by,

$$\text{EVII}(e) = \mathbb{E}_{\theta,z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{ u(a, \vartheta) \} \right\} - \max_{a \in \mathcal{A}} \mathbb{E}_\theta \{ u(a, \theta) \} \quad (2.14)$$

which is just the difference between the prior decision utility (Eq. 2.2) and the pre-posterior decision utility (Eq. 2.7).

The greater the uncertainty reduction provided by the measurement, the larger the **EVII**. So when there are multiple measurement options available, we can compare the improvements in decision making performance they provide to their costs, and determine which measurement (if any) is most economical for supporting our decision making. Fig. 2.7 illustrates this trade-off between the usefulness and cost of more precise measurement.

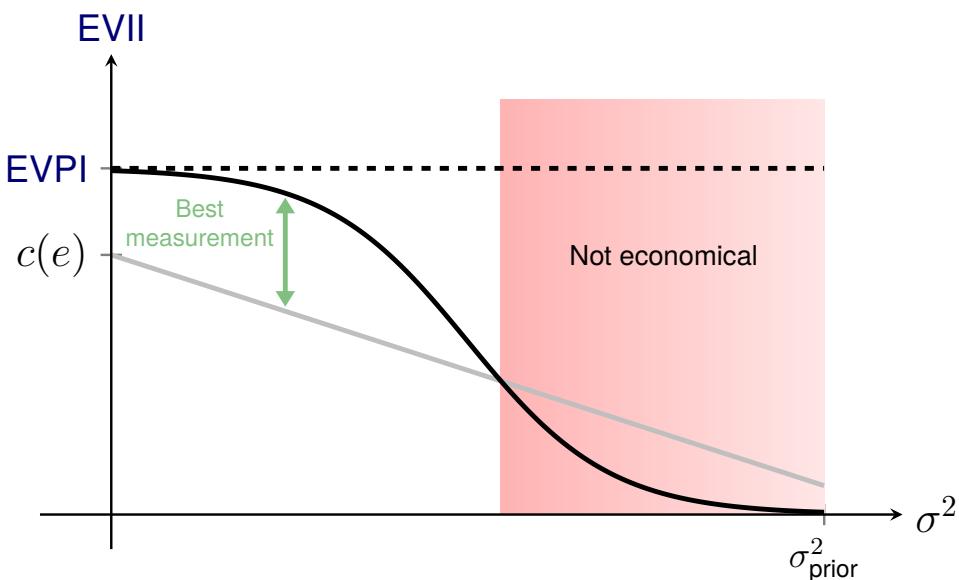


Fig. 2.7 Illustrative **EVII** vs. measurement cost curve

This figure shows a stylised relationship between the expected value of information (**EVII**, black line) and the cost of measurement (grey line) as measurement precision varies (given by the variance of the likelihood fn along the x-axis). The ‘best measurement’ (indicated in green) is the one which gives the highest net benefit (**EVII** minus cost).

²⁸ The previous perfect info. cases are just special cases of this, and provide an upper bound on this general value,

$$\mathbb{E}_{\theta,z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{ u(a, \vartheta) \} \right\} \leq \mathbb{E}_\theta \left\{ \max_{a \in \mathcal{A}} u(a, \theta) \right\}$$

²⁹ The **EVII** is also commonly referred to as the **Expected Value of Sample Information (EVSI)** (Keisler et al., 2014).

However we don't just need to compare different levels of reduction of the same uncertainty. As long as we can come up with a probabilistic model for the measurements, we can compare the benefit of reducing different uncertainties or combinations of them. When we're thinking about reducing multiple uncertainties in our problem we have to be careful, as the combination of uncertainty reductions can have a significant effect on the **Vol**. This is referred to as 'system effects' (Di Francesco, Langtry, et al., 2023). **Vol** values don't add. If two measurements provide complementary information then the **Vol** of the combination can be substantially greater. For example, when 'smart charging' an electric vehicle in a car park, knowing the state-of-charge of the vehicle may only be useful if we also know how long it'll be parked for, as only then can we know whether there's enough time to adjust the charging schedule. On the other hand, two measurements can contain redundant information, meaning the second provides little extra value. For example, supply/return water temperatures and heat meters provide similar information about a building's heating load.

2.E.5 Designing with energy usage data

Returning to the building electricity meter data we used in [2.E.2](#) to reduce uncertainty in the annual energy usage, let's now look at the value of that information for supporting our design decision. Fig. [2.8](#) compares the designs we would make with reduced load uncertainty to the initial design we would make with only the prior information, as well as the average cost of each design over the remaining uncertainty.

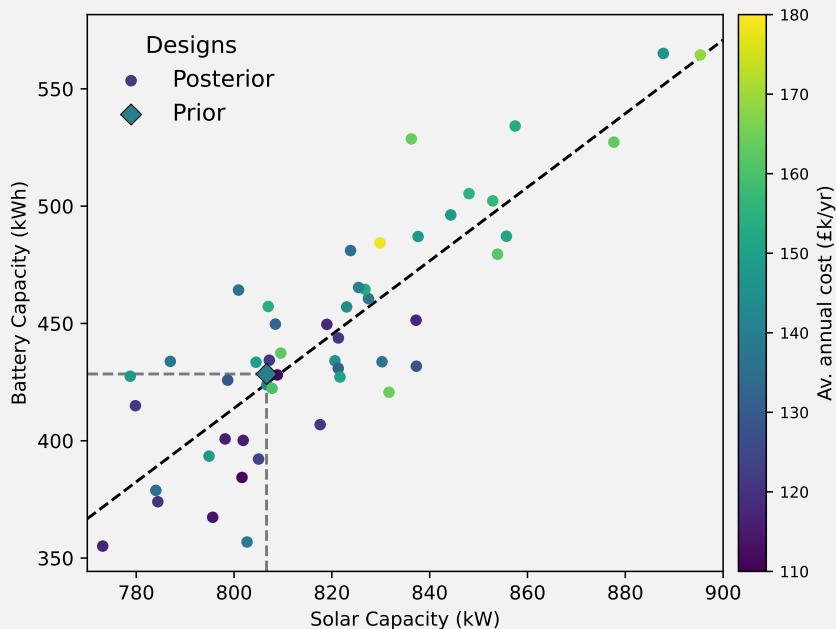


Fig. 2.8 Comparison of initial design and designs made with reduced load uncertainty.

As we commented last time, getting this information and reducing the load uncertainty leads us to significantly change our design choice, and we can end up making a design decision quite different from the initial choice we'd make (the prior design). Also, the expected cost of the system varies substantially ($\pm 25\%$) over the possible load measurements we could receive. This seems to suggest that the load uncertainty is having a significant impact on our decision making, as reducing it leads to large changes in both decision and cost.

But ultimately we're not actually interested in the choice we end up making. Instead, what we care about is how good a job we can do at solving our problem. The average cost of the designs made with reduced load uncertainty is £139.7k/yr. Meaning the **EVII** is just £0.2k/yr, or 0.2% of the prior cost. This is around half the **EVPI** value from [2.E.4](#) ([Table 2.3](#) below gives the comparison). This means that, even though the practical measurement removes around 75% of the uncertainty, it's only providing us with 50% of the benefit for decision making.

From this perspective, reducing load uncertainty in fact isn't particularly important to our decision making. It wouldn't be worth spending much money collecting this data, as the improvement it provides to our design decision is very small.

However, it does significantly reduce our uncertainty in how much our energy system will cost, which may be beneficial for financial planning and risk management. This is explored in more detail in [Chapter 5](#).

Case	From	Av. total cost of design(s) (£k/yr)	Vol (£k/yr)
No uncertainty reduction (prior)	Table 2.2 (SP soln.)	139.9	–
Imperfect information	2.E.5	139.7	0.2 (EVII)
Partial perfect information	2.E.4	139.4	0.5 (EVPI)
Perfect information	2.E.3	136.4	3.5 (EVPI)

Table 2.3 Summary of designs made with different levels of uncertainty reduction and associated **Vol** values.

*Remember that for the partial and imperfect information cases, only uncertainty in the building load was being reduced. We could potentially get somewhat closer to the **EVPI** value by also reducing the uncertainty in the battery cost and efficiency, for example, by negotiating with a manufacturer and doing our own tests on the batteries. But we could never reduce our uncertainty in the amount and pattern of solar generation.*

2.4.4 Information theoretic Vol

There is also an information theoretic view on **Vol**. It takes a purely statistical perspective, and defines the utility as the information content or statistical ‘surprise’ of the outcomes³⁰. This means that the expected utilities become entropies,

$$u(\theta) = -\log p(\theta) \implies \mathbb{E}_\theta\{u(\theta)\} = \sum_\theta (-p(\theta) \log p(\theta)) = H(\theta) \quad (2.15)$$

And so the **Vol** is the mutual information between the uncertainties and the measured data,

$$\text{Vol}(e) = \mathbb{E}_\theta\{u(\theta)\} - \mathbb{E}_{\theta,z}\{\mathbb{E}_\vartheta\{u(\vartheta|z)\}\} = H(\theta) - H(\theta|z) = I(\theta; z) \quad (2.16)$$

As this view of **Vol** is purely statistical, it doesn’t require a decision making model. This is both an advantage and a flaw. It means that we can calculate **Vol** values for measurements/uncertainty reduction without having to develop a potentially complicated model of how that uncertainty will be used to inform decision making. But as a result, we lose a lot of the interpretability of **Vol**. The units of **Vol** are now information ‘bits’, and we cannot know how useful that information will be for supporting our decision making, which is ultimately what we care about. This approach is most useful for comparing the uncertainty reduction provided by different measurements in complex systems³¹, and looking at the trade-off between the number of measurements made and total uncertainty reduction.

2.5 Framework extensions

The decisions that need to be made in real-world engineering problems have complexities that cannot be handled by the standard **Vol** framework that we’ve seen so far. This section presents two extensions to the framework that allow us to investigate the value of uncertainty reduction to improve decision making in more complex settings.

Firstly, ‘On-Policy’ **Vol**, which is a novel framework extension and contribution of this thesis, that allows the **Vol** to be studied in complex systems where direct optimisation is not possible. And secondly, the **Value of Optionality (VoO)** framework, which allows the benefit of retaining decision optionality as uncertainty reduces to be studied.

³⁰ ‘Surprise’ is a statistical measure of how surprising or unlikely an outcome is. The more unlikely an outcome, the more you learn (the more information you get relative to the prior) by observing it.

³¹ If the system is too simple the likelihood functions of the measurements already tell us what we need to know.

2.5.1 On-Policy Value of Information

So far, the **Vol** calculations have assumed that we have access to an accurate model of the system that we are able to optimise to find the decision/action that provides the highest expected utility, and also tell us that expected utility. However, many practical engineering problems are too complex for an accurate model of the system to be optimised directly. Instead, decisions are made using policies, such as Rule-based Control, Reinforcement Learning, or **Stochastic Programming**. These policies either use heuristics to guide their choice of, or search through, actions, or they simplify the system model and/or the statistical representation of the uncertainties³² and/or the optimisation objective (utility model), to provide an approximate model that is computationally manageable and can be optimised. As a result, a policy may not provide a sufficiently accurate³³ estimate of the expected utility to provide useful **Vol** values. I.e. the errors in the expected utility estimates may be larger than the **Vol**, and lead to extremely noisy **Vol** estimates.

To extend the standard **Vol** framework to allow decision making via policies, an additional step is added to the computations, which estimates the expected utility achieved by the actions selected by the policy using a more accurate model of the system, uncertainties, and utility.

A decision policy is defined as a mapping from a utility function describing a system, $u(a, \theta)$, and a distribution describing the uncertainties in the system, $\pi(\theta)$, to an action,

$$\mathcal{P}(u(a, \theta), \pi(\theta)) \rightarrow \alpha \in \mathcal{A} \quad (2.17)$$

As the utility function and set of possible actions are constant when we're looking at a given decision problem, we'll simplify the notation and refer to a policy applied to a given decision problem as $\mathcal{P}_u(\pi(\theta))$. The action selected by the policy therefore depends on the uncertainties in the system³⁴.

The **Vol** calculations proceed in the same way as before (compare to Eq. 2.14), but instead of selecting actions using a stochastic optimisation, they are now selected by applying the policy to the relevant distribution of uncertainties, and the expected utility

³² The most common example of this is the use of Monte Carlo approximations of the expected utility which use only a small number of samples to limit the computational cost of the model, but come at the expense of significant statistical error in the estimate.

³³ Or in some cases even estimate the expected utility at all.

³⁴ However we don't impose any conditions on how the policy uses the statistical information about the uncertainties, $\pi(\theta)$. The policy could be deterministic. For example, it could use only the expected (or maximum likelihood) values of the uncertainties from the distribution. In fact, this is what the **Expected Value Problem** does, and so it is a heuristic policy. *When policies don't make use of information about uncertainty, understanding how uncertainty reduction affects their ability to make decisions is, if anything, more important. The EVP, for example, cannot hedge its decisions, and so may be more reliant on good data.*

those actions achieve needs to be evaluated using the utility function/model.

$$\text{EVII}_{\mathcal{P}}(e) = \mathbb{E}_{\theta, z} \left\{ \mathbb{E}_{\vartheta|z} \left\{ u\left(\mathcal{P}_u(\pi(\theta|z)), \vartheta\right) \right\} \right\} - \mathbb{E}_{\theta} \left\{ u\left(\mathcal{P}_u(\pi(\theta)), \theta\right) \right\} \quad (2.18)$$

where the \mathcal{P} subscript indicates that this is the value of uncertainty reduction for supporting decisions made using policy \mathcal{P} .

This framework extension is called ‘On-Policy Value of Information’, as the decisions are made using a specific policy, and so the benefit of data collection/uncertainty reduction calculated is now dependent on the policy used. This broadens one of the main criticisms of the Vol framework, namely that Vol values do not generalise. This criticism is discussed further in Section 2.6. However, this does also open up the opportunity to study how different policies benefit from uncertainty reduction, and which are best able to exploit improved statistical information.

The key advantage of On-Policy Vol is that it allows us to study the value of information for complex, practical engineering problems where decision policies have to be used. For instance, problems with high-dimensional actions (i.e. where a large number of decisions must be made simultaneously), or those with tight computational limitations.

2.E.6 Using a decision policy

Throughout our example problem we’ve been using a Stochastic Program, specifically a linear scenario program, to perform our design optimisation and make decisions for us. This SP uses a simplified model of the energy system, it is linearised and assumes perfect foresight. Also, to keep the computation time manageable, when we solve the posterior decision problems we’ve only been using 10 samples from the posterior distributions of the uncertainties, leading to quite a noisy estimate of the expected utility.

So we are actually already using a policy to make our decisions. And there’s potentially quite a bit of error in our estimate of the expected utility from both the model and statistical simplification. We could create a more accurate simulator of our building energy system, and apply it to more samples from the posterior distributions to get a more accurate estimate of the expected utilities, and use this to do an On-Policy Vol calculation. However this is a bit too complicated for our simple example. You can see On-Policy Vol calculations for a more complex building energy system in Chapter 5.

2.5.2 Value of Optionality

Standard **VoI** calculations also assume that when decisions are made with and without data collection/uncertainty reduction, the set of possible actions is the same. However in many practical engineering problems, collecting data leads to a delay³⁵ before the decision is taken, and keeping decision options available through this delay is costly, even if those options do not end up being used.

We can adapt the **VoI** framework to study the effect that changing the set of available actions has on decision making as uncertainty is reduced. This is referred to as the **Value of Optionality (VoO)**.

If after data collection, the set of available actions is reduced from an initial set, \mathcal{A} , to a smaller set, $\mathcal{B} \subset \mathcal{A}$. Then the improvement in decision making that would be achieved by maintaining optionality, and keeping the original set of actions for the decision made with reduced uncertainty, referred to as the **Expected Value of Optionality (EVO)**, is given by,

$$\begin{aligned} \text{EVO}(\mathcal{B}, e) &= \mathbb{E}_{\theta, z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} - \mathbb{E}_{\theta, z} \left\{ \max_{a \in \mathcal{B}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} \\ &= \mathbb{E}_{\theta, z} \left\{ \max_{a \in \mathcal{A}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} - \max_{a \in \mathcal{B}} \mathbb{E}_{\vartheta|z} \{u(a, \vartheta)\} \right\} \end{aligned} \quad (2.19)$$

Notice that the inner term is zero unless we use the optionality and choose an action outside of the reduced set \mathcal{B} when data z is collected. This equation could also be expressed as a difference of **VoI** values.

Similarly to the **VoI**, the **VoO** tells us how much better we do on average at solving our problem by maintaining optionality and being able to choose from the full set of actions³⁶ when uncertainty has been reduced. By comparing the benefit optionality provides to decision making to the cost of keeping those options available, we can determine if maintaining optionality is worthwhile.

³⁵ This delay may form part of the cost of collecting data.

³⁶ We could also investigate the benefit of increasing the set of actions available to us.

2.E.7 Exploring optionality

For the design of our building energy system, we likely have a number of different options for decarbonising the energy usage. For example, instead of installing Li-ion battery storage we could instead install heat pumps, a hot water storage system, or mechanical ventilation with heat recovery (MVHR). We will likely go through an initial design phase and choose one of these approaches. After this, we'll do a more detailed design of the system, and gather some data about the building to help with this, for instance the historic electrical & heating loads and more precise data on the specs of our chosen system. However, the best decarbonisation approach depends on how much electricity and heat the building will use, and how our energy system performs. So after we've reduce these uncertainties we may find out a different approach than the one we initially chose would actually be better. But we're committed now.

We could perform a [VoO](#) calculation to see how much we could save by doing a preliminary design for all configurations of the energy system, gathering the building load data and tech specs from the equipment manufacturers, and then choosing the best approach after the uncertainties have been reduced. With this we could determine whether the extra time and effort required to provide this optionality is worthwhile, or if our initial design choice is good enough.

Again, doing this calculation is a bit complicated for our simple example. But a [VoO](#) calculation is performed in the context of energy storage technology selection for an industrial scale energy park in Chapter 6.

2.6 Limitations of framework

The key limitation of [Vol](#) analysis is that the [Vol](#) values calculated are only valid for the particular system model, statistical model, decision objective, and policy used. The framework does not provide any guarantees that the conclusions about which and how much data should be collected will generalise to other systems, or even with changes to the current model and its assumptions. So care needs to be taken when interpreting and communicating the insights provided by the [Vol](#), being clear that the conclusions only apply to how data collection is used to support the specific decision made within the assumptions and limitations of the model. Of course this is a very common limitation of engineering analyses, particularly optimisation modelling. Generalisation guarantees are difficult to provide, and at some point judgement is required as to whether a model is a good enough

representation of the practical engineering problem. However, **Vol** can be particularly sensitive to the model setup. This is because **Vol** values depend on the utility values at the optima, which can move significantly and unexpectedly as the model changes. Seemingly innocuous changes to the model can cause large changes in **Vol**, for example if they lead to a dominant action which causes the **Vol** to drop to zero. Adding new uncertainties into the model can also affect the **Vol**, even if those uncertainties are not measured and reduced.

This generalisation limitation can be overcome by performing sensitivity analyses to test whether the conclusions drawn remain valid as the model setup and assumptions change. However, this is very computationally expensive, and it's not possible to test all model variations that could be relevant. Particularly as performing **Vol** calculations is already highly computationally expensive.

The second main limitation is the computational cost of calculating the **Vol**. In **Vol** analysis we are testing how the decision optimisation changes as different data values are observed. Outside of very simple analytic cases, this requires us to sample possible data values and repeat the decision optimisation a large number of times to get a good **MC** estimate of how well we can do on average when making decisions with data. So computing a single **Vol** value requires likely hundreds of optimisations to be performed. For practically useful problems, these optimisations are themselves computationally expensive. This means that either expensive compute resources³⁷ are needed, the scope of problems considered needs to be limited, or significant simplifications need to be made to the model, which can affect the accuracy of the results. For the **Vol** calculations in Chapters 5 & 6, a supercomputer³⁸ was needed to run the experiments in an achievable time.

To be able to perform the decision optimisations, we also need to be able to develop a decision-making model of the engineering system, be that a system model that can be optimised or a suitable decision policy. For complex, real-world engineering systems this modelling can be extremely challenging and require significant domain expertise. The **Vol** framework also requires that we describe outcomes with a single utility value. This means it is not compatible with multi-objective decision making and cannot account for subjective factors of decision making that often occur in engineering design. The framework is most suited to automated or model-based decision making.

Vol calculations investigate the impact that data has on the decision model being tested. So, when interpreting the **Vol** values that we receive from the analysis and making judgements about whether data should be collected, we are assuming that the decision

³⁷ The optimisations needed for **Vol** calculations are embarrassingly parallel. So calculations can be performed quickly if enough compute is available.

³⁸ The experiments were run using the Cambridge Service for Data Driven Discovery (CSD3, www.csd3.cam.ac.uk).

model provides a sufficiently good representation of the physical engineering system we care about. Specifically, we assume that the usefulness of data in the model is the same as in the real system. While the **Vol** framework studies the quality of data in detail, it does not have any mechanism for accounting for the quality of the model. The On-Policy **Vol** methodology presented in Section 2.5.1 provides some route to addressing this limitation, as the decision model (policy) and evaluation model (utility) are separated, allowing a more computationally expensive, and hopefully more accurate model to be used to evaluate the effect of the decisions on the system. But it still requires the assumption that this more accurate model is a good enough representation of the real system.

This limitation of model quality is inherent to analyses concerned with design, as the system does not yet exist, and so cannot be measured to validate the model. Expert judgement and experience of other systems is needed to assess whether a particular model is good enough to provide trustworthy **Vol** values. However, in some cases we can justify that the model used for **Vol** is, in a certain sense, exact. This is because **Vol** is concerned with how uncertainty affects decision making. If we have access to the decision model that real-world practitioners use to make their decisions, then using that model in the calculations tells us exactly how uncertainty affects those decisions. For example, many energy systems are designed using linear stochastic program models. The analysis in Chapter 5 tells us exactly how a designer using such a model is affected by building load uncertainty, and how useful load data is to them. So in an industry setting, **Vol** analyses can use the actual decision making models. This is, provided they are computationally feasible. Using a simplified model would lead to the same quality limitations.

The model quality includes not only the accuracy of the physics of the system, but also the statistical models of the uncertainties and data. **Vol** calculations can be sensitive to these statistical models, but must assume that they are correct³⁹. This includes aspects such as the resolution of the data in the model, and correlations between uncertain variables.

We collect data for many reasons, and data has many uses. However, **Vol** calculations only quantify the value of data collection to support the specific decision studied. So if that same data is useful for other purposes not captured by the model, this extra value will be missed by the calculated **Vol**. It's difficult, for example, to describe how useful having a historic record of some measurement of a system might be to inform future decision making, such as maintenance. However, having the **Vol** value helps us make these finer judgements about whether data is worthwhile collecting. With knowledge of the value of that data for supporting the known and modellable decisions, and the cost of collecting the

³⁹ However, precisely defining what it is for a predictive statistical model to be inaccurate, and identifying how this inaccuracy affects the **Vol**, is not an easy task. After all, a prior distribution represents our best knowledge about an unknown quantity. Considering whether this knowledge is wrong can lead to pathological and unhelpful analyses.

data, we only need to judge whether we think the extra value we might get from the data justifies that difference between the known benefit and the cost. And it may turn out that the data is already worth collecting on the basis of the known decisions.

Finally, the standard [Vol](#) framework assumes that decisions are made in a risk-neutral way. This comes from the optimisation objective being the expected/average utility. However many real-world decision makers are risk averse, particularly those making large financial investments to decarbonise energy. The framework can easily be adapted to account for risk aversion in decision making by including it the decision objective. This allows us to study how uncertainty reduction affects decision risk, as is done in Chapter [6](#).

2.E.8 Understanding and overcoming our limitations

In our example problem, we found that collecting electricity meter data to reduce the uncertainty in the building load provided little benefit for improving our design decision of sizing the solar-battery system. But the calculations we've done only apply to the specific building energy model we've used, the probabilistic models of the uncertainties, and our decision objective. All of these contain simplifications and assumptions which might not be the same for other buildings. So to make a more general statement about the value of energy usage data for supporting building energy system design, we'd need to see whether our conclusion stays the same when we adapt our model to the properties of other buildings. For example, domestic buildings will have different load distributions and energy sale prices.

To make our example computationally achievable, we've had to make some significant simplifications to the building energy model. We've linearised the physics and assumed perfect foresight. Potentially if we used a more realistic model we might get quite different cost values, which could alter the [Vol](#). There are also some uncertainties we haven't included in the model, for example uncertainty in the cost of solar panels, which might also affect the [Vol](#).

To keep the computation time down, we had to use quite a small number of samples from the posterior distributions. So there might be quite significant statistical error in our estimate of the [Vol](#) values. To test this, we should perform some more samples and see whether the [Vol](#) values change significantly, or whether the estimates have already converged.

Appendix 2.A Simple building energy system model for example problem

For our example problem of designing (specifically sizing) a solar-battery system for a commercial building, we'll consider a simplified model of the building energy system. This system has four key components: the building electrical load which needs to be met (we'll assume the heating is electrified), the solar panels which generate electricity, the battery storage system which can store and release electricity, and the grid from which we can buy and sell electricity. Fig. 2.9 illustrates the energy flows in this energy system model.

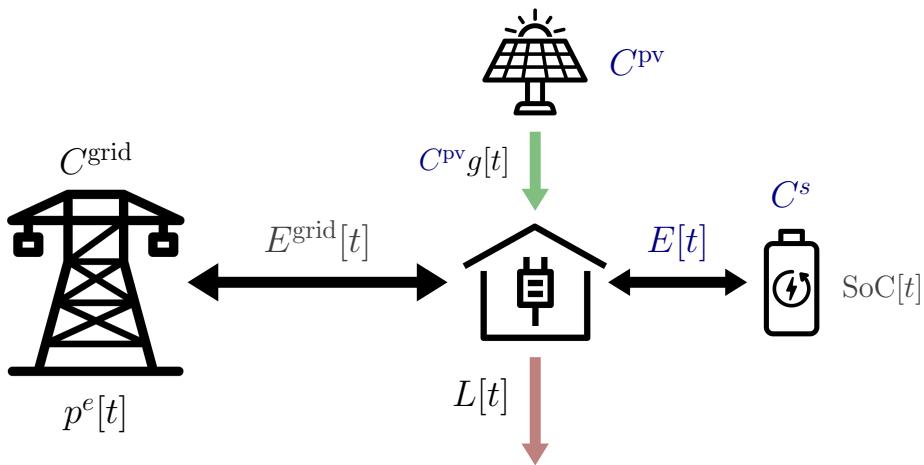


Fig. 2.9 Illustration of energy flows in example building energy system.

2.A.1 Energy model

We can create a linear model of the behaviour of this energy system by assuming that the only physics we need to satisfy is the conservation of energy, and that the properties of the solar panels and battery storage system (e.g. efficiency, power limits, etc.) are fixed/constant. We will treat the building load, generation potential of the solar panels, and price of grid electricity as exogenous, but potentially uncertain.

Our goal is to design the system to minimise the total cost of meeting the building's energy demands. This is made up of the cost of installing the solar panels and battery storage, the cost of buying electricity from the grid, and the revenue from selling excess electricity back to the grid. To reduce computation, we'll simulate one year of operation of the system and compute the annualised total cost. It's assumed that when we sell electricity back to the grid, we receive some fraction of the market price we would buy at, and that there is a constraint on the maximum power that can be drawn from the grid, i.e. a grid connection capacity.

This linearised model of the energy system can be described using a Linear Program (**LP**). This is given by Eq. 2.20, with the parameters of described in Table 2.4. A software implementation of this **LP** is available [here](#).

The model can be run either in simulation or design mode. In simulation mode, capacities for the solar and storage are specified, and the operation of the system (scheduling of the battery storage) is optimised, meaning the model returns the annualised total cost of the specified design if it were operated optimally. In design mode, the capacities of the solar and storage are included as decision variables in the optimisation, meaning the model simultaneously optimises the operation and sizing of the system, returning the lowest possible annualised total cost and the system design that achieves it.

For the example experiments we will use historic building load monitoring data from a university building in Cambridge⁴⁰, as well as solar generation and electricity pricing data from the area. This is the same data that is used for the investigation in Chapter 5. In fact, Chapter 5 is just a more complex version of this example problem. The other parameter values used in the experiments are given in Table 2.5.

$$\min \quad p^{\text{PV}} C^{\text{PV}} + \sum_m \rho_m \left(p_m^s C^s + \sum_t p_e[t] \left([E_m^{\text{grid}}[t]]^+ + \kappa [E_m^{\text{grid}}[t]]^- \right) \right) \quad (2.20)$$

over $C^{\text{PV}}, C^s, E_m[t], \text{SoC}_m[t+1] \quad \forall m, t$

subject to $\text{SoC}_m[t+1] = \text{SoC}_m[t] + \sqrt{\eta_m} [E_m[t]]^+ - 1/\sqrt{\eta_m} [E_m[t]]^- \quad (2.20\text{a})$

$$-P^{\max} \Delta t \leq E_{i,m}[t] \leq P^{\max} \Delta t \quad (2.20\text{b})$$

$$(1-\nu) C^s \leq \text{SoC}_m[t+1] \leq C^s \quad (2.20\text{c})$$

$$P^{\max} = \delta C^s \quad (2.20\text{d})$$

$$\text{SoC}_m[0] = \text{SoC}^0 C^s \quad (2.20\text{e})$$

$$E_m^{\text{grid}}[t] = L_m[t] - C^{\text{PV}} g_m^{\text{PV}}[t] + E_m[t] \quad (2.20\text{f})$$

$$E_m^{\text{grid}}[t] \leq C^{\text{grid}} \Delta t \quad (2.20\text{g})$$

for all $m \in [0, M-1], t \in [0, T-1]$

⁴⁰ Specifically, we use the load data from Building 0 in the dataset from Chapter 5 for this simple example problem.

Parameter	Units	Description
Decision variables		
C^{pv}	kWp	Installed solar PV capacity
C^s	kWh	Installed energy capacity of battery storage
$E_m[t]$	kWh	Energy flow <i>into</i> battery storage at time t in scenario m
$\text{SoC}_m[t]$	kWh	State-of-charge of battery storage at time t in scenario m
Derived variables		
$E_m^{\text{grid}}[t]$	kWh	Net energy <i>drawn from</i> electricity grid at time t in scenario m
P^{\max}	kW	Power capacity of battery storage
Sampled parameters		
ρ_m	—	Probability of scenario m
$L_m[t]$	kWh	Building electrical load at time t in scenario m
$g_m^{\text{pv}}[t]$	kWh/kWp	Solar PV power generation potential at time t in scenario m
η_m	—	Round-trip efficiency of battery storage in scenario m
p_m^s	£/kWh/yr	Annualized capacity cost [†] of battery storage in scenario m
Known parameters		
Δt	hrs	Time step of simulation data
δ	kW/kWh	Discharge ratio of battery storage (power capacity/energy capacity)
ν	—	Depth-of-discharge of battery storage
SoC^0	—	Initial state-of-charge of battery storage (fraction of capacity)
C^{grid}	kW	Grid connection capacity
p^{pv}	£/kWp/yr	Annualized capacity cost [†] of solar PV
$p^e[t]$	£/kWh	Price of grid electricity at time t
κ	—	Electricity sell price as fraction of buy price
Indices		
t	—	Time step in modelled operation
m	—	Scenario number

Table 2.4 Description of variables & parameters of Stochastic Program, Eq. 2.20.

[†] Capital cost per energy/power capacity per year of lifetime

Parameter	Units	Value	References
Simulation duration, T	Hours	8760	
Simulation time step, Δt	Hours	1	
Discharge ratio, δ	kWp/kWh	2	(Kebede et al., 2022)
Depth-of-discharge, ν	—	0.9	(IRENA, 2017)
Initial state-of-charge, SoC ⁰	—	0.75	
Grid capacity, C^{grid}	kW	500	
Annualised cost of solar PV	£/kWp/yr	75	(Ramasamy et al., 2023)
Fractional sell price, κ	—	0.5	
<i>Default values of uncertain parameters</i>			
Solar year	—	2017	
Load year	—	2017	
Average building load	kW	100	
Battery efficiency, η	—	0.95	
Battery cost, p^s	£/kWh/yr	70	(Forbes Home, 2024)

Table 2.5 Parameter values for example building energy system model.

2.A.2 Probabilistic model of uncertainties

In our example we'll look at uncertainties in the building load, solar generation, and battery storage. Where uncertainties in these factors are considered, we'll use the probabilistic models given in Table 2.6, otherwise we'll treat them as fixed and use the values given in Table 2.5.

The uncertainty in the building load is represented using two parameters: the load year which represents the uncertainty in the shape and timing of the load profile, and the average building load. Years of load data are sampled uniformly from the available dataset.

Parameter	Distribution	Parameters	Units
Solar year	Discrete uniform	{2012, ..., 2017}	—
Load year	Discrete uniform	{2012, ..., 2017}	—
Average building load	Truncated Normal (cut-off at $\pm 2\sigma$)	$\mu=100, \sigma=10$	MW
Battery efficiency, η	Truncated Normal (cut-off at $\pm 2\sigma$)	$\mu=0.95, \sigma=0.05$	—
Battery cost, p^s	Truncated Normal (cut-off at $\pm 2\sigma$)	$\mu=70, \sigma=5$	£/kWh/yr

Table 2.6 Distributions of uncertainties.

2.A.3 A brief tirade on constraints in statistical optimisation

The mathematics of optimisation do not sit very comfortably with a probabilistic view of the world. This is especially true when it comes to constraints. Mathematical constraints are written as strict laws that the problem must adhere to. But when we come to solve optimisation problems numerically, Lagrangian optimisation treats these constraints in the same way as the objective (Boyd and Vandenberghe, 2004), and allows us to violate them, though our solvers try not to. In the real world there are some strict constraints, we have to obey the laws of physics. But often we impose practical engineering constraints on our system that are somewhat artificial.

The problem comes when we try to model uncertainties in our system that interact with these constraints. If we don't account for the full value range of the uncertainties it's possible that we will end up making decisions that will violate the constraints in some possible realisations. But, if we try to make decisions that will satisfy the constraints in every possible outcome we'll end up being far too conservative. The issue is that in our optimisation we can't differentiate between the hard constraints we *have* to obey, and the softer constraints where there is potentially some flexibility.

For instance, in our example problem the deterministic LP chooses a system design that is infeasible when we start accounting for uncertainty in the solar generation. This is because the design puts in a large solar generation capacity, and in years where the sun is more intense, the solar power ends up exceeding the grid connection capacity. From a mathematical perspective it's difficult to know what to do in this case. Our optimisation simply fails. In reality, there is likely to just be some cost associated with exceeding the grid connection capacity, either a charge from the grid operator (Economics, 2022) or a cost associated with having to disconnect the solar panels and causing some damage to the system.

Because we have no clear way of dealing with infeasibility and failed optimisations, we instead need to model our energy systems in such a way that our uncertainties cannot lead to violation of the constraints. Where we have soft constraints, we should instead model the costs associated with violating them and include these in the objective, as is done for grid connection capacity in Chapter 5. Ultimately we need to consider how our system will actually behave when things go wrong, and account for the associated costs in our decision making. A classic example of this in the energy systems literature is constraining energy systems to have some specified loss-of-load probability. This constraint is rather arbitrary, and is very hard to guarantee. Instead we should be concerned with the cost of not satisfying the load, as we can then account for this cost trade-off when we design the energy system.

3

Applying Vol to energy systems

Max Langtry, Chaoqun Zhuang, Rebecca Ward, Nikolas Makasis, Monika J. Kreitmair, Zack Xuereb Conti, Domenic Di Francesco, and Ruchi Choudhary (Nov. 2024). “Rationalising Data Collection for Supporting Decision Making in Building Energy Systems Using Value of Information Analysis”. In: *Journal of Building Performance Simulation*, pp. 1–17. ISSN: 1940-1493. doi: [10.1080/19401493.2024.2423827](https://doi.org/10.1080/19401493.2024.2423827). URL: <https://doi.org/10.1080/19401493.2024.2423827>

This chapter has been published as the journal article above, as well as a corresponding conference paper (**Langtry**, Zhuang, et al., 2023). The example decision problems in this chapter were developed in collaboration with the co-authors of the article, who are experts on the different energy systems studied. They contributed to the development of the energy system models for each example, and provided a small amount of the text describing those models that appears in this chapter. Other than these contributions, all of the code, analysis, and text in this chapter is my own work.

All code and data used to perform the experiments in this chapter is available at <https://github.com/EECi/VOI-for-Building-Energy>.

Having introduced the **VOI** methodology and how it works in the previous chapter, this chapter demonstrates how **VOI** can be used to understand data collection requirements in building energy systems. It explores three example decision problems in building energy systems: air source heat pump maintenance scheduling, ventilation scheduling for indoor air quality, and ground-source heat pump system design. In each a different aspect of how **VOI** can be used to support decision making is illustrated. The examples show how **VOI** can be used prioritise, rationalise, and justify expenditure on data collection in building energy systems, and avoid wastage on low insight data.

3.1 Introduction

As explored in the [Introduction](#), data can be collected to support decision making in the design and operation of building energy systems by reducing uncertainties, allowing more informed and effective decisions to be taken (Molina-Solana et al., [2017](#)). However, data collection is costly. The cost of data collection must be weighed against the benefits it provides to decision making to determine whether it is economically worthwhile.

Accounting for uncertainties during planning and the use of data collection to support decision making has been standard in the building energy systems field for some time (Mavromatidis et al., [2018a](#); Mobaraki et al., [2022](#); Rysanek and Choudhary, [2013](#); Tian, Heo, et al., [2018](#)). However, few studies that recommend collecting data discuss its cost. Studies that do quantify data collection costs frequently report only the sensor cost (K. Han and J. Zhang, [2020](#); Zhai and Salazar, [2020](#)), omitting several components of the overall cost of data collection, such as the costs of digital infrastructure, data processing, maintenance, and quality assurance. Very few studies in the building energy literature have examined the economic benefit of data collection, and quantified the value provided by data collection. There may exist significant unidentified wastage in building systems in the form of low insight monitoring. This wastage will likely grow if data collection strategies are not rationalised as building digitisation continues to expand.

The **Vol** framework introduced in Chapter 2 provides a principled, numerical methodology for quantifying the value of data collection in building energy systems. This allows expenditure on data collection and monitoring systems to be justified, prioritised, and rationalised, and for cost optimal data collection strategies to be determined. So far, no study has used **Vol** to quantify the benefit that uncertainty reduction from *practical* data collection would provide to decision making in building energy systems.

This chapter demonstrates how the **Vol** methodology can be used to investigate data collection requirements for building energy systems. It presents three example decision problems:

- Smart meters to support maintenance scheduling for **Air Source Heat Pumps**
- Occupancy monitoring systems to support ventilation scheduling for indoor air quality in offices
- Ground thermal testing to support the design of **Ground Source Heat Pump** based heating systems for apartment blocks

Each showing a different decision making context: management, operation, and design respectively.

In each example, **Vol** is used to study how data collection can improve decision making, and the insights that it can provide are discussed. The examples illustrate different aspects of **Vol**: representation of complex systems, sensitivity to model assumptions, and determination of optimal measurement precision.

Sections 3.2 to 3.4 present the three example decision problems in turn. In each section, the model of the decision problem is described, **Vol** analysis is performed to study a different aspect of the benefit of data collection, and the insights it provides into the usefulness of that data are discussed. Finally, conclusions are drawn in Section 3.5.

3.2 Condition measurement for optimising maintenance scheduling for air source heat pumps

Air Source Heat Pumps (ASHPs) provide significant advantages to decarbonising the heating of buildings through their ability to exploit ambient heat in the environment to achieve high Coefficients of Performance (COPs), reducing the direct energy input required to heat a space, and so reducing the embodied carbon emissions of heating. However, through usage, the performance of ASHPs degrades, reducing the COPs they can achieve. Maintenance activities can be undertaken to address this performance degradation and improve the COPs achieved by the heat pumps, reducing the electricity consumed and operational costs. However, regular maintenance is costly.

When deciding how frequently to maintain ASHP units, asset owners seek to trade-off the cost of maintenance activities with the benefits they provide in reduced electricity consumption cost to minimise the total cost of operating the ASHP asset. However, the performance of degrading ASHPs depends on several uncertain factors, including the degradation rate, the base Seasonal Performance Factor (SPF), and the performance improvement provided by maintenance. Additionally, at the time maintenance is scheduled, the cost of electricity and the heating load of the building are unknown. Therefore, the maintenance scheduling decision must be made under these uncertainties.

Smart meter data can be used to better estimate the performance of ASHP units, allowing for the scheduling of maintenance to be optimised. However, installing and maintaining smart meters adds additional cost to the operation of the ASHP units. Therefore, asset owners will raise the question, “Does installing a smart meter on an ASHP unit reduce the overall operating costs by allowing for optimised maintenance scheduling?”.

A case involving a research building at the University of Cambridge, which is equipped with 30 identical ASHP units, is examined to assess the economic viability of installing a centralised electricity smart meter to assist maintenance scheduling optimisation. The asset owner selects the number of evenly spaced maintenance activities undertaken per year, $N_m \in \{0, \dots, 12\}$, as to minimise the expected operating cost of the heating system (Eq. 3.10).

The annual energy consumed by the ASHP units is given by,

$$E = \frac{L_H}{\text{SPF}} \quad (3.1)$$

where L_H is the heating load of the building for the coming year. This heating load is assumed to be Gaussian distributed, with mean and standard deviation determined from historic building metering data,

$$L_H \sim \mathcal{N}(L_H, \mu = 12.6, \sigma = 1.36) \text{ GWh/year} \quad (3.2)$$

The total cost of electricity consumed to meet the building heating load is given by,

$$C_e = p_e E \quad (3.3)$$

where p_e is the price of electricity for the coming year, which is assumed Gaussian with mean and standard deviation calculated using consumer electricity cost data across UK regions from 2022 (DESNZ, 2023a),

$$p_e \sim \mathcal{N}(p_e, \mu = 32.6, \sigma = 1.6) \text{ p/kWh} \quad (3.4)$$

The annual Seasonal Performance Factor (SPF) of the ASHPs, the average COP over the heating season, is given by,

$$\text{SPF} = \text{SPF}'(1 - \alpha)(1 + \beta) \quad (3.5)$$

where SPF' is the base heat pump SPF, which is not known at the time of installation. Using data from existing ASHP units (Nouvel et al., 2015), it is assumed to be distributed as,

$$\text{SPF}' \sim \mathcal{N}(\text{SPF}', \mu = 2.9, \sigma = 0.167) \quad (3.6)$$

α is the performance degradation factor of the ASHPs, which is modelled as being distributed as a truncated Normal with mean 0.01 and standard deviation 0.25, as it is a non-negative parameter,

$$\alpha \sim \mathcal{N}(\alpha, \mu = 1e-2, \sigma = 0.25 : \alpha \geq 0) \quad (3.7)$$

β models the performance improvement provided by maintenance activities, and is given by,

$$\beta = \frac{\beta_a N_m^\gamma}{\beta_b + N_m^\gamma} (1 + \varepsilon) \quad (3.8)$$

the parameters β_a , β_b , γ , are empirical parameters with assumed values of 0.05, 2.5, and 1.4 respectively (Griffith et al., 2008). The parameter ε models uncertainty in the performance

improvement, and is taken to be distributed as,

$$\varepsilon \sim \mathcal{N}(\varepsilon, \mu = 0, \sigma = 0.1) \quad (3.9)$$

The cost per activity of performing maintenance on all 30 ASHP units, C_m , is taken to be £18,000 (Daikin, 2022). So, the total cost of operating the ASHPs (comprised of electricity and maintenance costs) is given by,

$$C_{\text{total}} = C_e + C_m N_m \quad (3.10)$$

The described stochastic decision problem of optimally scheduling ASHP maintenance is represented in influence diagram form in Fig. 3.1, where grey arrows represent measurements. This diagrammatic representation provides a far clearer, more concise, and readily extendable description of the decision task within this complex engineering system.

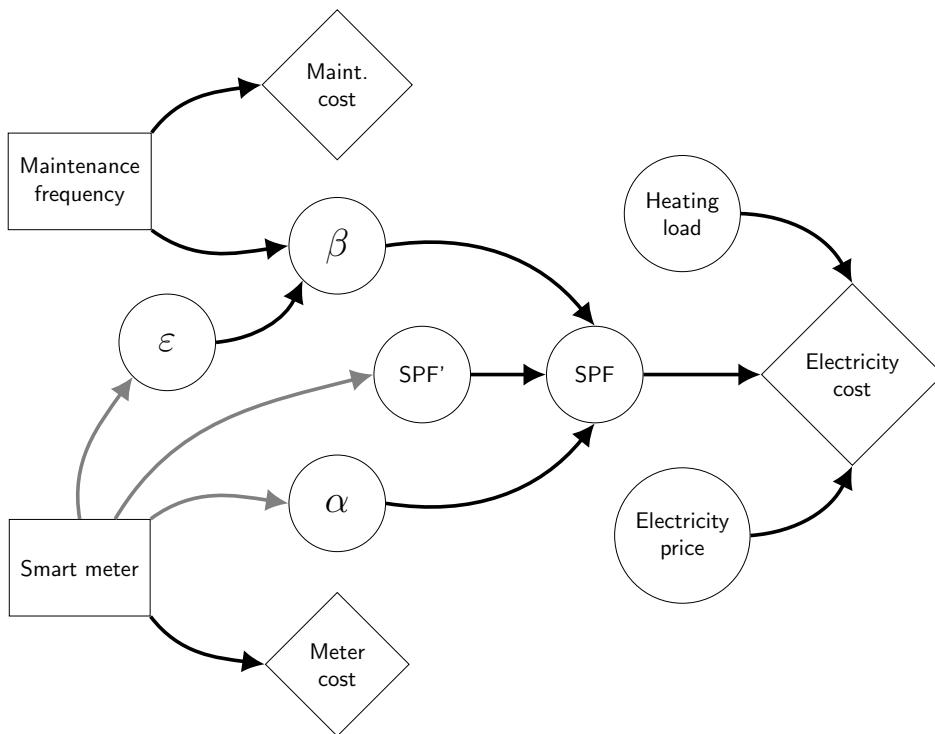


Fig. 3.1 Influence diagram representation of ASHP maintenance scheduling decision problem

Solving the Prior decision problem⁴¹, the optimal maintenance frequency is found to be 2 activities per year, leading to an expected operating cost (Eq. 3.10) of £1,876,300/year for the building heating system, of which £36,000/year (2%) is spent on maintenance. Fig. 3.2 plots the probability distribution of operational costs for the system when maintained 2 times a year. It demonstrates the significance of the uncertainties in the building energy system on the planning problem, as the cost of operating the heating system can vary from £1m/year to £4m/year, and so highlights the importance of accounting for those uncertainties during planning.

Solving the Pre-Posterior decision problem⁴¹ with perfect information, i.e. planning maintenance with perfect knowledge of the uncertain parameters in the system, achieves an expected cost of £1,874,640/year. Therefore the Expected Value of Perfect Information (EVPI), as defined in Section 2.4, is £1,660/year.

Fig. 3.3 shows the distribution of optimal maintenance frequency when the true values of the uncertain system parameters are known. In slightly over half of cases, perfect information about the system does not change the maintenance scheduling decision, i.e. the true solution is the same as the prior solution, indicated by the black column. However the information provides value, as in many cases the performance degradation of the ASHPs is less severe, with less frequent maintenance providing lower operational cost. Monitoring information allows the system operator to avoid the cost of unnecessary maintenance activities in these cases.

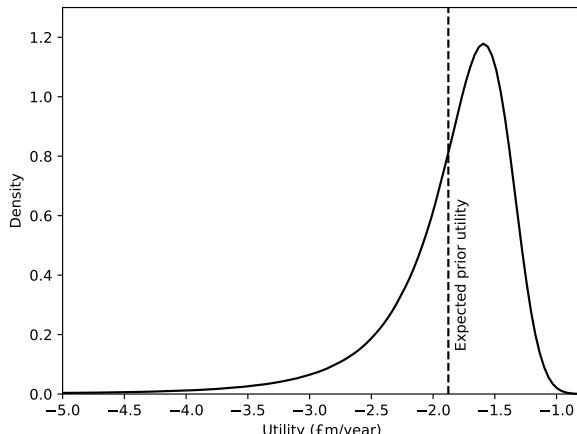


Fig. 3.2 Distribution of utilities achieved by optimal prior action, $N_m^* = 2$. Dashed line indicates mean of distribution

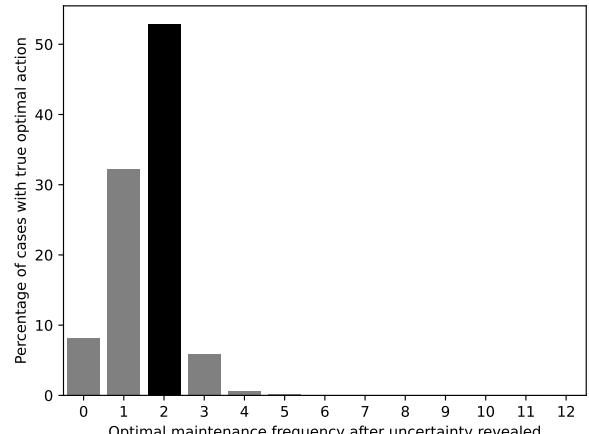


Fig. 3.3 Distribution of true optimal maintenance frequency, i.e. best action once uncertainty has been revealed

⁴¹ Refer back to Section 2.3 for definition.

The annualised cost of the electricity smart meter used to monitor the ASHP units ('meter cost' in Fig. 3.1) is estimated to be £70/year (Daikin, 2022). Therefore, the installation of smart meters capable of perfectly measuring ASHP performance, degradation, and building load, and the dynamic scheduling of maintenance activities using the actual electricity price, would lead to a net economic benefit, Vol minus information cost, of £1,590/year to the asset owner from reduced operational costs. In this way, the asset owner can justify their investment in smart meters. The value of information provided by smart meters in this scenario is small compared to the total operational costs, only 0.06%. However, the use of smart meters to optimise maintenance scheduling can provide additional benefits, such as extending the lifespan of the ASHP units and decreasing the probability of malfunctions. These cost savings, although not accounted for in these calculations due to simplification, could be significant.

3.3 Building occupancy measurement for real-time ventilation scheduling to improve indoor air quality

In mechanically ventilated office spaces, building managers must schedule ventilation system settings to ensure sufficient indoor air quality for the occupants. Adequate ventilation is required to prevent the transmission of airborne infectious diseases such as the SARS-CoV-2 virus (COVID-19), which is both damaging to the health of the occupants and costly for the tenant of the office space through lost productivity. However, operating ventilation at an unnecessarily high rate can impact occupant thermal comfort, lead to excessive carbon emissions, as well as excessive operational cost of the ventilation system through the additional heating/cooling demand required to condition air intake. The risk of viral transmission, and thus the appropriate ventilation setting, is highly dependent on the number of occupants in the office space. However, in the absence of occupant monitoring and dynamic ventilation control systems, building managers must schedule ventilation system settings without knowledge of the exact occupancy level of the space. At the time of deciding ventilation scheduling, the occupancy of the ventilated space over the operation period is uncertain. This uncertainty is particularly relevant in light of recent trends towards 'work from home', which reduces the predictability of office occupancy.

This raises the question, "Would it be worth installing a smart occupancy monitoring and ventilation control system in an office space to improve ventilation scheduling?", i.e. would the economic benefit of improved ventilation control be greater than the cost of installing such a smart monitoring and control system?

A simple model of indoor air quality in a typical office space is considered. Said office space is taken to have a maximum occupancy of 100 people, a floor area of 1,000 m², and a ceiling height of 2.4 m. There are five available ventilation settings: 1, 3, 6, 12, and 20 air changes per hour (ACH). The ventilation system is assumed to have a fan of specific power 1.9 W/l/s, P_{fan} , operating at 60% efficiency, η , for 10 hours per day, T . At an electricity unit price of 32.6 p/kWh, p_e , the cost of operating the ventilation system⁴² at each setting is calculated as,

$$C_{\text{vent}} = \text{ACH} \times V \times \frac{P_{\text{fan}}}{\eta} \times T \times p_e \quad (3.11)$$

where V is the volume of the office in litres, T is expressed in seconds and p_e in £/Wh.

A model of the probability of viral infection for an individual in an indoor space as described in (de Oliveira et al., 2021) is used, which is also available in web application form at [airborne.cam](#) (Gkantzas et al., 2021). It is assumed that the base prevalence of infection amongst the occupants is that of the general UK population in February of 2023, 2.18% (ONS, 2023), and that infection of an individual leads to 3 days of sick leave, which taking the median daily salary for full-time employees in 2022, costs the tenant £128/day (ONS, 2022) in lost productivity. It is further assumed that any occupants are present in the office space for the whole 8 hour work day, that time coupling effects between days can be neglected, i.e. that the model for a single day is representative, and so illness costs are calculated using the expected number of infections for the given occupancy. The prior distribution of occupancy is taken to be discrete uniform in the interval 0 to 100 inclusive.

The stochastic decision problem is thus to select the ventilation system setting which minimises the sum of the ventilation system operation (electricity) cost, and the cost of lost productivity due to illness to the tenant, subject to uncertainty in the occupancy level of the space.

$$C_{\text{total}} = C_{\text{vent}}(\text{ACH}) + n \times p_{\text{infection}}(n, \text{ACH}) \quad (3.12)$$

where n is the number of occupants in the office.

This decision problem can be described via the influence diagram provided in Fig. 3.4.

Solving the Prior decision problem, the optimal ventilation rate is found to be 12 air changes per hour, leading to an expected overall cost of £138/day for the tenant. Fig. 3.5 shows the distribution of utilities for each available ventilation rate under the prior uncertainty. Increasing the ventilation rate increases the base cost of all outcomes, indicated by the dashed line showing the minimum cost for each curve, but reduces the risk of large numbers of infections occurring when the office has high occupancy, limiting the maximum cost.

⁴² The cost of heating/cooling associated with ventilation is neglected for simplicity, but can be readily added to the model.

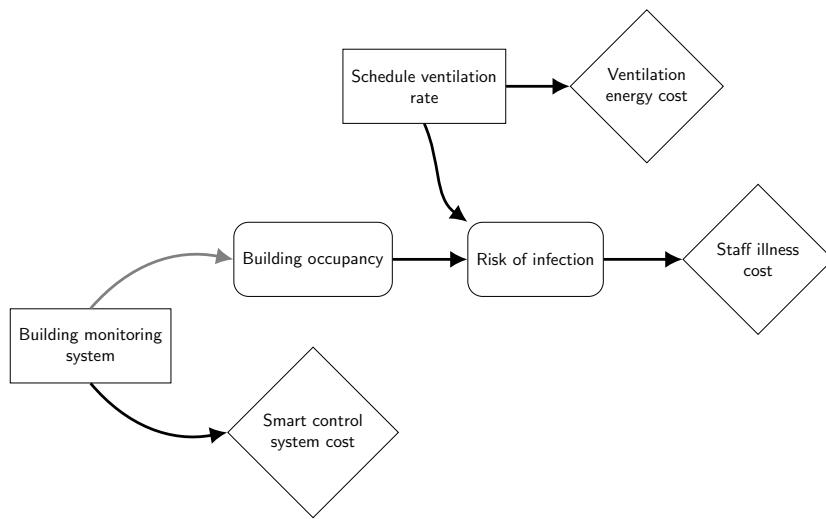


Fig. 3.4 Influence diagram representation of building ventilation scheduling decision problem

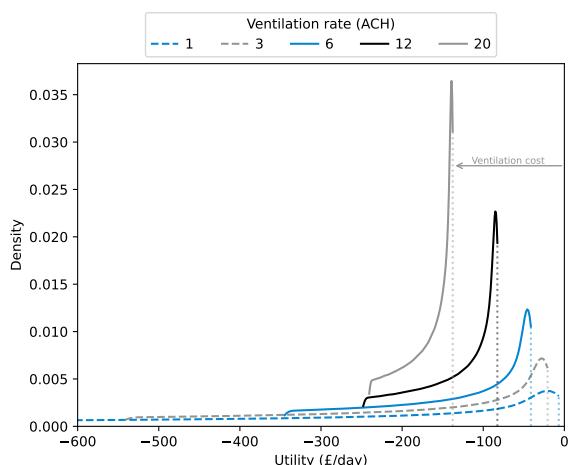


Fig. 3.5 Distribution of utilities achieved by each ventilation rate under prior uncertainty

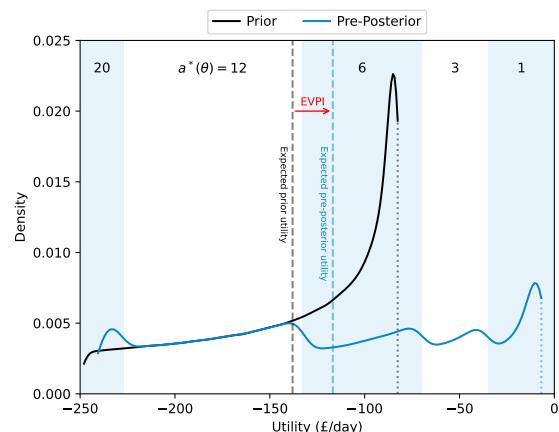


Fig. 3.6 Distribution of utilities achieved by prior and pre-posterior decisions

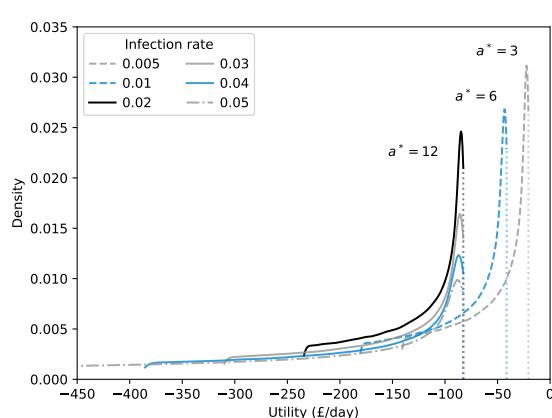


Fig. 3.7 Distributions of utilities for prior solution with varying assumed infection rate

The utility distribution for the prior optimal ventilation rate of 12 ACH in the base case is indicated by the black curve in Figures 3.5 to 3.7.

Solving the Pre-Posterior decision problem with perfect information, i.e. dynamically scheduling ventilation with perfect knowledge of occupancy from a monitoring system, achieves an expected cost of £117/day. Therefore, the **EVPI** is £21/day. Hence, over a 20 year operational lifetime, the smart monitoring and control system could save the tenant up to £850,000, or 15% of the total operating cost. Fig. 3.6 plots the distribution of utilities achieved by the prior and pre-posterior decisions. For the prior, the same decision of 12 ACH is used in all cases as no measurement is taken. Whereas, for the pre-posterior, the ventilation rate is set using information on the occupancy level in each case. The blue regions indicate the regime of cases where each ventilation rate is selected by the pre-posterior decision, i.e. where it is the lowest cost choice. The pre-posterior density is similar to the prior in the region of high costs, but is much more uniform and extends substantially further into the low cost region, indicating that the **Vol** is primarily derived from the ability to reduce ventilation rate and save electricity cost when office occupancy is low.

Building managers can use the computed **EVPI** to support their decision making about whether investment in smart occupancy monitoring and ventilation control systems are worthwhile. If a proposed smart monitoring and control system is expected to have an installation and maintenance cost over the 20 year lifetime of more than £850,000, then this investment will not be beneficial for the context of ventilation scheduling. A cheaper but less precise method of estimating occupancy, such as a desk booking system, may be a more suitable strategy, or it may be most cost effective to forego occupancy measurement altogether and use a fixed ventilation rate.

Sensitivity analysis

The problem formulation considered assumes a variety of properties about the office being ventilated. However, commercial building stocks are diverse, and the actual value of these properties varies between buildings. A sensitivity analysis is performed to investigate whether occupancy monitoring and dynamic ventilation scheduling remains valuable across different sized offices. The **Vol** analysis is repeated for office models with 5 to 25 m² of floor area per person, maintaining a maximum occupancy of 100, and keeping all other properties as before. Table 3.1 provides the optimal ventilation rate without measurement and the **EVPI** for each size of office considered. The **EVPI** is lowest in the extreme cases where extreme ventilation rates are optimal both on average (for the prior problem) for the majority of occupancy outcomes. For instance, in the smallest office, the highest available ventilation rate of 20 ACH is the optimal prior decision, and the optimal decision for 53% of occupancy outcomes. For all office sizes tested, the **EVPI** remains significant, 10% or

more of the prior expected cost, suggesting that occupancy monitoring systems are likely beneficial for offices of any size.

Floor area per person (m^2)	5	10	15	20	25
Prior optimal ventilation rate (ACH)	20	12	6	3	3
EVPI (£/day)	14	21	16	19	14

Table 3.1 Sensitivity of prior solution and **EVPI** to office size

The prevalence of viral illness in the general population has a significant impact on the risk of infection spreading in an office environment, and so should be taken into account by a building manager when setting ventilation rates, and as a result when determining what information is required to support that decision. The sensitivity of the prior decision and **EVPI** is tested for infection rates from 0.5% to 5% (ONS, 2023). Table 3.2 provides the results of this analysis, and Fig. 3.7 plots the distribution of utilities for the prior solutions at each infection rate. As infection rate increases, the risk of infection at a given occupancy level rises, causing a widening of the utility distribution, and so greater risk for the decision maker. Additionally, the distribution of true optimal ventilation rates flattens. Hence, for more occupancy outcomes a different ventilation rate to the prior is optimal, and the benefit of that improved decision increases. Therefore, the greater the infection rate, the higher the **Vol**, and the more valuable occupancy monitoring becomes. The sharp increase in **EVPI** between infection rates of 1% and 2% highlights the importance of using sensitivity analysis to determine whether conclusions about the economic benefits of data collection derived from **Vol** analysis remain valid as assumptions about the studied building energy system are varied. Alternatively, assumed parameters can be modelled as uncertainties and included within the system model.

Infection rate (%)	0.5	1	2	3	4	5
Prior optimal ventilation rate (ACH)	3	6	12	12	12	12
EVPI (£/day)	8	11	22	21	24	30

Table 3.2 Sensitivity of prior solution and **EVPI** to infection prevalence in the population

3.4 Ground conductivity measurement for optimising bore-hole design in residential ground source heat pump heat supply systems

Ground Source Heat Pump (GSHP) systems use the ground as a source and sink of heat to provide cooling and heating for buildings in a highly energy efficient way (Aresti et al., 2018). As these systems exchange heat with the ground, their performance (characterised by the COPs they achieve) is influenced significantly by the geological properties of the ground with which they exchange heat. In the design of such GSHP heating & cooling supply systems, it is desired to match the capacity of the GSHP system to energy demand of the building as to minimise the overall cost of operating the supply system over its lifetime. The overall operating cost is composed of the capital cost of constructing the system, and the operational costs, which are primarily the cost of the electricity required to meet the building energy demands. Under-specification of the systems leads to greater electricity usage from less efficient and so higher cost auxiliary heating systems, whilst over-specification of the system results in unnecessary capital cost.

At the time of system design, the thermal properties of the ground are typically not known precisely, as existing geological survey data provides an uncertain estimate of the ground properties in the site location (Dalla Santa et al., 2020). However, the system designer has the option to commission thermal tests at the site location prior to designing the GSHP system. Whilst these tests reduce the uncertainty in the ground thermal properties, they are time-consuming and incur additional costs. As a result, these tests are often not conducted, and generally available information on materials and location is used to provide uncertain estimates of ground thermal properties which are used to design the GSHP system. This therefore poses the following question to the designer, “Would commissioning thermal tests reduce the overall lifetime cost of the GSHP heat supply system by improving the matching of the designed system capacity to the building load?”.

A simplified GSHP system design task for a residential building heat supply system is considered. In this design task, the designer must select the length of boreholes, L_{bh} , to be drilled for the ground heat exchange. The available length choices are 110 to 190 m, in 5 m increments. The capital cost of borehole drilling is taken to be 70 £/m/borehole.

It is assumed that the effective ground thermal conductivity, λ_{ground} , is the only uncertain geological parameter, and that it is Normally distributed with mean 1.94 W/mK, and standard

deviation 0.31 W/mK,

$$\lambda_{\text{ground}} \sim \mathcal{N}(\mu = 1.94, \sigma = 0.31) \text{ W/mK} \quad (3.13)$$

which is the average thermal conductivity over the range of borehole depths considered, and covers typical uncertainty ranges given the heterogeneity present in soils (Busby, 2018, 2011; Loveridge et al., 2013).

The designed GSHP system, consisting of 12 boreholes, must supply heat to a small apartment block of 10 flats over a 50 year operational lifetime. The building is assumed to have a typical heating demand distribution for its type, and a simplified load profile is synthesised based on demand values for the UK (Mitchell and Natarajan, 2020), and using historic weather data for London. This synthetic load profile, $E_{\text{load}}(t)$, consumes 116MWh/year, corresponding to a 13.2 kW mean heating load for the building, with a peak load power of 30.5 kW. The data can be found in the [GitHub repository](#).

The model of geothermal borehole operation presented in (Lamarche and Beauchamp, 2007) is used to determine the scheduling of energy extracted from the ground in each time instance, $E_g(t)$, with the borehole fluid temperature, T_{fluid} , conservatively constrained to be within the range 5 degC to 35 degC,

$$5 \leq T_{\text{fluid}}(t) \leq 35 \quad (3.14)$$

In this model, T_{fluid} is a function of the power extracted from the ground, $E_g(t)$, the ground thermal conductivity, λ_{ground} , the borehole length, L_{bh} , and the other assumed ground condition parameters.

Given the fluid temperature schedule determined, the instantaneous COP of the GSHP system is computed using the following empirical relationship from (Kensa, 2014),

$$\text{COP}(t) = 4.0279 + 0.1319 \cdot T_{\text{fluid}}(t) \quad (3.15)$$

The energy that is provided by the GSHP system to the building is then given by,

$$E_{\text{GSHP}}(t) = \frac{E_g(t)}{1 - 1/\text{COP}(t)} \quad (3.16)$$

If the GSHP system is unable to meet the building load in any time instance, the remaining unsatisfied load is provided by an auxiliary heat supply system with a COP of 1.

$$E_{\text{load}}(t) = E_{\text{GSHP}}(t) + E_{\text{aux}}(t) \quad (3.17)$$

The total electricity consumption of the combined heat supply system is therefore given by,

$$e_{\text{total}} = \sum_t \left(\frac{E_{\text{GSHP}}(t)}{\text{COP}(t)} + E_{\text{aux}}(t) \right) \quad (3.18)$$

The cost of electricity used by the residential building is taken to be 32.6 p/kWh.

The stochastic decision problem of designing borehole lengths as to minimise the expected lifetime cost of the heating supply system is represented as an influence diagram in Fig. 3.8.

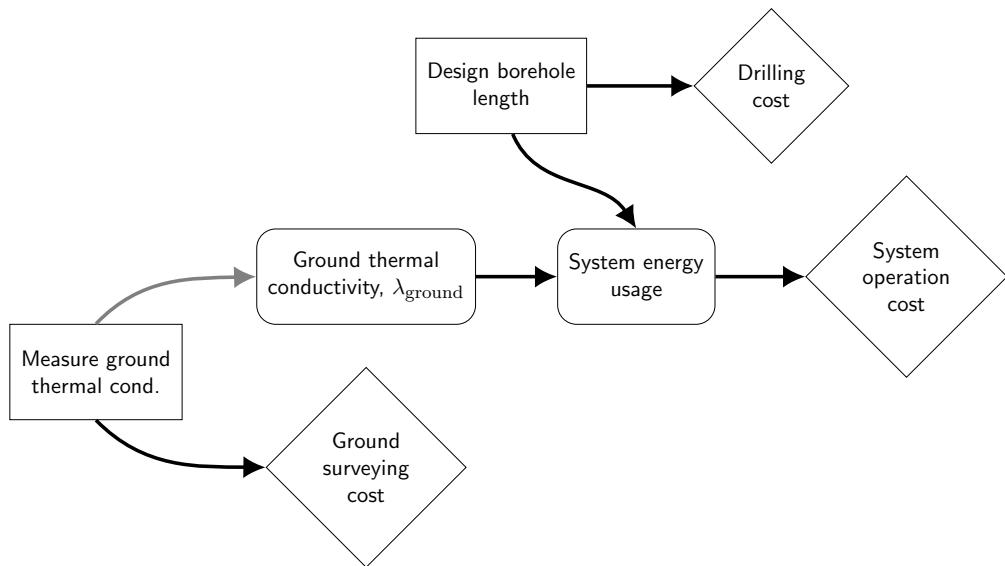


Fig. 3.8 Influence diagram representation of GSHP heat supply system design decision problem

Solving the Prior decision problem, the optimal borehole length is found to be 155 m, leading to an expected overall lifetime cost of £819,100 for the heating system.

Various technologies exist for measuring ground thermal properties. Table 3.3 presents the cost and experimental uncertainty of thermal probe tests and Thermal Response Tests (TRT), which are commonly used for ground thermal conductivity measurement. The extended TRT measurement is a hypothesised base case for the results achievable with TRT methods. The likelihood of obtaining a measurement z for each test given a true underlying ground thermal conductivity of λ is modelled using a Gaussian distribution as,

$$f_e(z|\lambda) \sim \mathcal{N} \left(z, 0, \frac{\nu_e}{2} \lambda \right) \quad (3.19)$$

where ν_e is the decimal experimental uncertainty of the test.

Method	Uncertainty	Cost (£)	References
Thermal probe (in situ)	25%	187	(King et al., 2012), (Rentals, 2023)
Thermal probe (lab test)	17%	1,800	(Low et al., 2013), (Laboratory, 2024)
Thermal Response Test (TRT)	10%	5,000	(Choi, Choudhary, et al., 2021), (Spitler et al., 2000), (Tang and Nowamooz, 2019)
Extended TRT	5%	10,000	–

Table 3.3 Uncertainty and cost of ground thermal conductivity measurement tests

The **Expected Value of Imperfect Information (EVII)** is computed for each available test, which is compared to the associated cost to determine the net benefit to decision making provided. Fig. 3.9 plots the results. All tests provide significant net benefit to the heating system design task, indicating that site specific ground thermal testing should be conducted prior to the design of **GSHP** based heating systems. The TRT provides the greatest net added value to the decision maker, and so using this **Vol** analysis, the designer of this energy system can justify and optimise their expenditure on ground testing to support their decision making. This demonstrates that additional uncertainty reduction through more data collection or more precise measurement does not always provide sufficient improvements to decision making to warrant its cost.

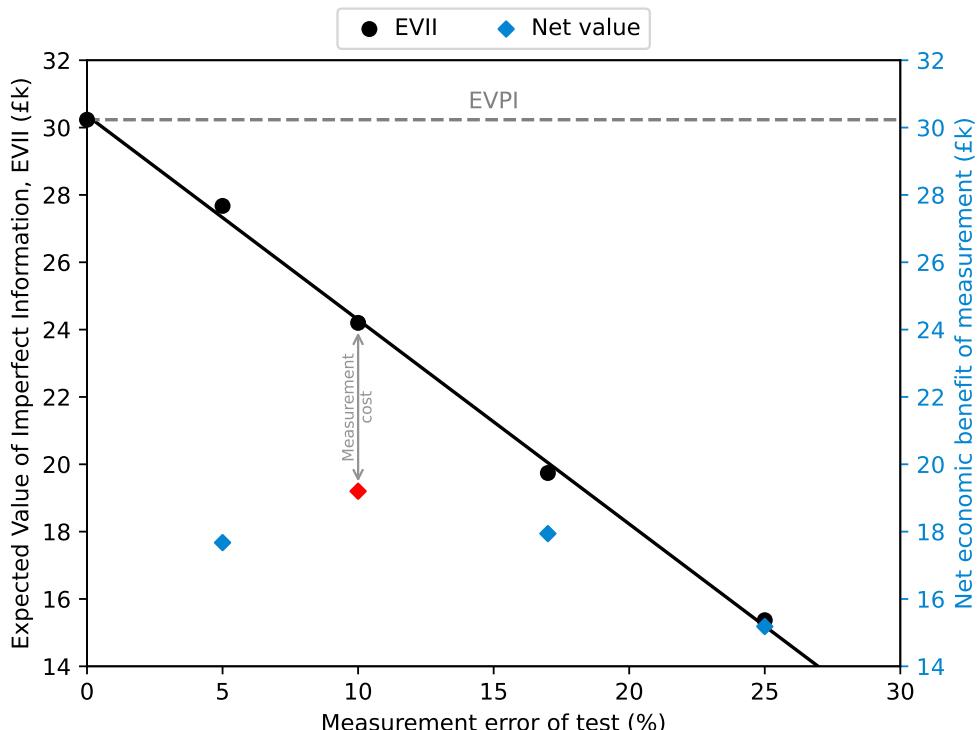


Fig. 3.9 Expected Value of Imperfect Information (EVII) and net economic benefit (EVII minus measurement cost) of imprecise ground thermal conductivity measurements

3.5 Conclusions

This chapter demonstrated how [Vol](#) can be used to quantify the benefit of data collection to support decision making in building energy systems, and as a decision support tool for evaluating data collection strategies. Three example decision problems covering the management, operation, and design of building energy systems were studied: heat pump maintenance scheduling, office ventilation scheduling, and ground source heat pump system design.

Installing smart meters in [ASHPs](#) for dynamic maintenance scheduling was found to be a worthwhile investment. However, improvements in decision making provided were extremely limited, with perfect knowledge of heat pump condition leading to only a 0.06% reduction in operational costs on average. Occupancy monitoring systems enabled significant cost savings when scheduling ventilation to manage indoor air quality and infection risk in office spaces, reducing the total cost to the tenant by 10.2-16.4%. Occupancy information was found to be more valuable in mid-sized offices and when infection is more prevalent, e.g. in the winter. When designing a [GSHP](#) based heating system for a block of flats, ground thermal tests to reduce uncertainty in thermal conductivity were found to be beneficial to support system design. A Thermal Response Test (TRT) was determined to be the optimal data collection strategy, enabling net savings of £19k on the lifetime cost of the heating system.

In each example problem a different capability of the [Vol](#) framework was demonstrated: the ability to simply and clearly represent and analyse decisions in complex engineering systems, the identification of system characteristics for which data collection is worthwhile, and the determination of optimal data collection strategies when trading-off measurement precision versus cost.

The examples in this chapter show that [Value of Information](#) analysis gives a clear and principled methodology for addressing the current research gap of quantifying the benefit of data collection to support decision making in the building energy systems literature. It provides a route to rationalise data collection strategies and avoid resource wastage on low insight data as building digitisation progresses. However, these examples use significantly simplified models of how decisions are made in building energy systems, and don't go into much depth about how the uncertainty reduction that the data provides affects the decision making process.

4

Data requirements of load forecasting for model predictive control

Max Langtry, Vijja Wichitwechkarn, Rebecca Ward, Chaoqun Zhuang, Monika J. Kreitmair, Nikolas Makasis, Zack Xuereb Conti, and Ruchi Choudhary (Oct. 2024). “Impact of Data for Forecasting on Performance of Model Predictive Control in Buildings with Smart Energy Storage”. In: *Energy and Buildings* 320, p. 114605. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2024.114605](https://doi.org/10.1016/j.enbuild.2024.114605). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824007217>

This chapter has been published as the journal article above. The co-authors of this article developed the code for the simple neural models used in the experiments. They also performed the data similarity metric and change-point analyses, detailed in Appendices 4.D & 4.E, and the data feature and online training experiments in Section 4.3.3, providing a small amount of the text describing these analyses. Other than these contributions, all of the code, analysis, and text in this chapter is my own work.

All code and data used to perform the experiments in this chapter is available at https://github.com/EECi/Annex_37.

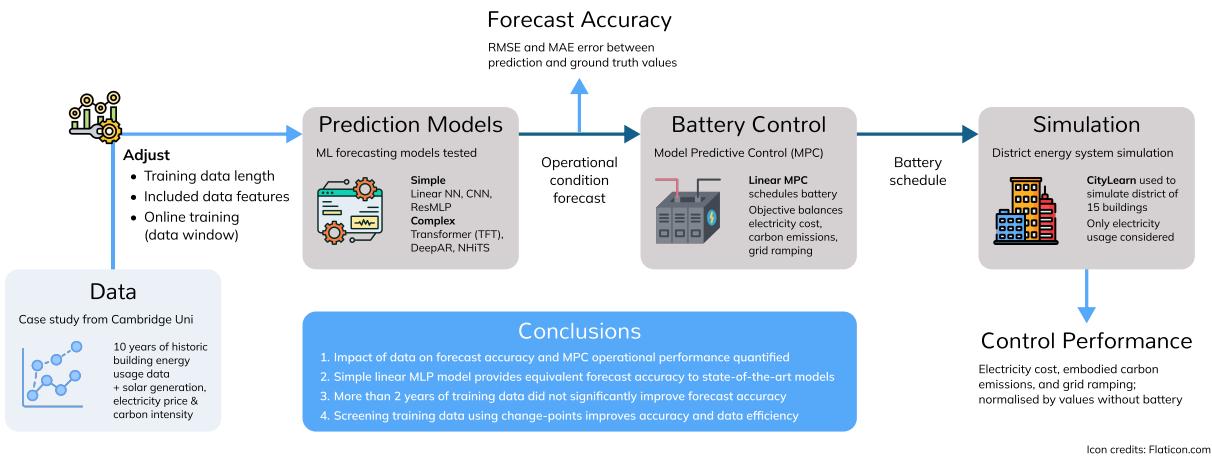


Fig. 4.1 Graphical abstract of Chapter 4.

Data is needed to develop forecasting models for building behaviour that are used for smart control of the building energy systems to reduce the cost and carbon emissions of building operation. However, this data is costly to both collect and exploit. Determining the most cost effective data usage strategies and setting data collection requirements for these systems requires understanding of how data affects forecast accuracy, and the resulting impact on the operational performance of the building control. This chapter investigates how the available data affects the forecasting accuracy of machine learning prediction models, and how the accuracy of these prediction models affects the performance of **Model Predictive Control** in a simulated multi-building energy system. The data requirements of state-of-the-art and simple textbook machine learning models are compared, and measures to improve the data efficiency of these models are investigated. Specifically, reuse of prediction models, reduction of training data duration, reduction of model data features, and online model training.

4.1 Introduction

Distributed generation and storage technologies, most commonly solar panels and batteries, are used to reduce the cost and or operational emissions building energy systems (Amini Toosi et al., 2022; O'Shaughnessey and Shah, 2021; X. Zhu et al., 2023). 'Smart' operation of these energy storage systems can reduce the impact the building's energy usage has on the electrical grid, by allowing energy flexibility through demand-side management and demand response (Vazquez-Canteli, Henze, et al., 2020). This is particularly important in light of the expected electrification of heating demands (Leibowicz et al., 2018), and the peak load issues this will cause. The effectiveness of these systems, quantified by operational performance metrics of the building energy system such as total electricity cost, incurred carbon emissions, and measures of grid impact, is determined by the ability of the storage control scheme to arbitrage energy and alter the timing of net energy usage to meet these operational goals. As a result, lots of research effort has been put into developing performant storage control strategies in recent years (Drgoňa et al., 2020; Kathirgamanathan et al., 2021; Z. Wang and Hong, 2020).

Model Predictive Control (**MPC**) is one of the two leading control methodologies in the current building energy systems literature, alongside Reinforcement Learning (**RL**). Its use in smart energy storage systems in buildings has been widely studied (Deng et al., 2023; D. Lee et al., 2020; Souto, 2018; Touretzky and Baldea, 2014; Zhou et al., 2021), and it has been found to provide substantial operational performance improvements compared to Rule Based Control (**RBC**) (Erfani, Xingji Yu, et al., 2021; D. Lee et al., 2020; Oldewurtel et al., 2012), the main technology used in existing building systems. It has also been shown to provide comparable performance to **RL** based control (Zhan, Lei, and Chong, 2023). Review papers (Farrokhifar et al., 2021; Thieblemont et al., 2017; Jun Wang et al., 2022) provide a thorough overview of the applications of **MPC** to building energy systems with distributed generation and storage.

MPC requires a model of the system dynamics, which estimates the next state of the system given; (a) the current system state, (b) current operational conditions, and (c) the applied control actions. Using this dynamics model and forecasts of the operational conditions over a given planning horizon, it predicts the operation of the system over that planning horizon. Optimization techniques are then used to determine the set of control actions that optimize a specified objective over the planning horizon. The performance of **MPC** therefore depends on: the accuracy of the system dynamics model, the accuracy of the operational condition forecasts, the ability of the optimization method to identify near-optimal control actions, and the match between the objective over the planning horizon and the global operational goals of energy management for the system. This chapter

focuses on the accuracy of operational condition forecasts, and its resulting impact on the operational performance of MPC. (Raisch, Langtry, et al., 2025) explores how data affects the accuracy of the system dynamics model, building on previous work on the topic by the main author (Raisch, Krug, et al., 2025).

4.1.1 Forecasting models for MPC

Lots of different time series forecasting methodologies have been studied for the predicting operational conditions for building energy management (Ying Sun et al., 2020b), however recent research has focused on machine learning based methods (Chou and Tran, 2018; Dai, Dilsiz, et al., 2023; C. Fan et al., 2019; Rahman et al., 2018; L. Zhang et al., 2021) due to their promising performance. Successes in other fields have prompted the investigation of large-scale, high-complexity model architectures (Choi and S. Lee, 2023; Dai, Niyogi, et al., 2023). However, there is doubt about whether such complex models are appropriate for use in the context of MPC for building energy systems (Bünning, Huber, et al., 2022; Bünning, Pfister, et al., 2021; Zeng et al., 2023), due to the computational and data availability limitations in practical systems. So both simple and high-complexity, state-of-the-art machine learning forecasting models are investigated.

4.1.2 Impact of data on forecast accuracy and MPC performance

As machine learning methods are purely data-driven, ‘black box’ prediction models, achieving accurate forecasts requires the availability of training data that is representative of the building energy system for which the model will be used. However, acquiring representative training data is both challenging and costly. The costs of acquiring data include not only the capital costs of installing monitoring systems, but also the costs of digital infrastructure, data processing, system maintenance, and quality assurance required to support data collection. For the case of developing forecasting models for MPC considered in this study, there are additional project costs associated with delaying the installation of the battery system while data is gathered. (Motegi et al., 2003) estimates the capital cost of electricity and gas smart-metering for a university campus to be \$0.27/m², plus an additional \$0.11/m² for maintenance and supporting IT systems. However, this considers only the cost of collecting data, and neglects the significant costs of training and deploying machine learning models (Strubell et al., 2020), from both the computing and expertise required. Whilst the importance of data availability and the impact of data on forecast accuracy are widely acknowledged in the existing literature (Choi and S. Lee, 2023; Kathirgamanathan et al., 2021; D. Lee et al., 2020; Jun Wang et al., 2022; Zhan and Chong, 2021), few works

study the role of data in enabling good operational performance for MPC in building energy systems.

Determining cost optimal data collection strategies to support the development of forecasting models for MPC in buildings requires understanding of the trade-off between the quantity of data & data features used for model training, and its associated costs, and the operational performance achieved by the controller. This impact of data for forecasting on MPC operational performance can be considered in two stages, by firstly studying the relationship between data and forecast accuracy of the resulting models, and then the sensitivity of the controller operational performance to the accuracy of forecasts.

Only a single previous study has investigated the impact of data on the accuracy of forecasting models for building energy management. This work, (Choi and S. Lee, 2023), compares the prediction accuracy of deep learning architectures for forecasting thermal loads and building zone temperatures over varying training dataset sizes. It finds that increasing the training data length does not always improve prediction accuracy due to the strong seasonality of building energy behaviours, but that the addition of data with good similarity to the test dataset into the training dataset greatly improves forecast accuracy. A limitation of this study is that it analyses only a single year of thermal load and zone temperature data, synthesised from a building energy model and historical weather data, meaning that a limited range of training dataset sizes, from 3 to 9 months, is considered. As a result, training data preceding the test data by one year, which due to seasonality is likely to have the highest similarity and so greatest value, is unavailable, meaning the study of the benefit of additional training data is incomplete.

The impact of forecast accuracy on MPC operational performance has been studied to a limited extent. (Erfani, Xingji Yu, et al., 2021) compares the use of various classical and machine learning forecasting models in a common MPC framework, quantifying both the forecast accuracy and resulting operational performance of MPC. However, the models studied all achieve comparable forecast accuracies and operational performances, meaning limited insight can be gained into the relationship between the two. Further, as comparison is made between model types, these results cannot be used to assess the benefit of forecast improvement for any individual model. (Oldewurtel et al., 2012) shows that the use of more accurate, external weather forecasts improves MPC operational performance, but does not quantify the forecast accuracies for comparison. In (Bartolucci et al., 2019), the prediction accuracy and corresponding operational performance of MPC for two forecasts of electrical load with different levels of synthetic noise are quantified. Finally, (Enríquez et al., 2016) provides the most complete study, investigating the variation of operational

performance with the noise amplitude of synthetic forecasts of temperature and solar irradiance. However, as the forecast accuracy of the synthetic predictions is not quantified, these results cannot be compared to the performance of practical prediction models and used to assess the benefits of forecast accuracy improvements.

The direct impact of training data length on operational performance is studied in (Savadkoohi et al., 2023). It computes the operational performance achieved by a neural network based building thermal controller as the size of the training dataset varies, and finds that negligible performance improvements are obtained when using more than 8 months of training data. Whilst a comprehensive quantification of the relationship between training data volume and operational performance achieved by the specific control scheme studied is provided, the results are specific to the atypical controller architecture used, which does not include the explicit forecasting of operational variables used in typical MPC schemes. Additionally, other aspects of data, such as the inclusion of additional data variables, and the reuse of data from existing buildings, are not considered.

Questions of the impact of data on MPC performance in buildings, and the optimal data collection strategies to support model development, have been explored in the System Identification field (Balali et al., 2023; Erfani, Jafarinejad, et al., 2023; Zhan and Chong, 2021; Zhan, Lei, Jin, et al., 2022; Zhan, Quintana, et al., 2022; Q. Zhang et al., 2023) in the context of developing accurate system dynamics models for MPC, termed ‘control-oriented models’. Studies have investigated the data requirements of different modelling approaches (Balali et al., 2023; Zhan and Chong, 2021; Q. Zhang et al., 2023), the impact of data resolution on model accuracy (Erfani, Jafarinejad, et al., 2023), the impact of model prediction accuracy on operational performance (Zhan, Lei, Jin, et al., 2022), and cost optimal data collection strategies to support model development (Zhan, Quintana, et al., 2022).

4.1.3 Research objectives & contributions

There has been very limited study of the impact of data on the prediction accuracy of forecasting models for building energy management. Additionally, there has been no study of the trade-off between the quantity of data & data features used for model training, and the operational performance of MPC for battery scheduling in systems with distributed generation and storage. Understanding of this trade-off is necessary to properly prioritise expenditure on data collection for smart energy storage systems. This chapter addresses this research gap by quantifying the impacts of data on both the prediction accuracy and operational performance of MPC using simple and high-complexity, state-of-the-art machine learning based prediction models. A simulated multi-building energy system with distributed solar generation and battery storage using historic building load measurement data is used as a case study. The main research objectives are to:

- Compare the performance of simple and state-of-the-art machine learning models with regards to prediction accuracy, model generalisation, and data efficiency;
- Investigate the trade-off between data and forecast accuracy for the following data efficiency measures: reuse of prediction models, reduction of training data durations, reduction of model data features, and online model training;
- Propose strategies for improving prediction performance when selecting models for reuse and selecting data to exclude when reducing training data durations; and
- Quantify the relationship between forecast accuracy and resulting operational performance of MPC.

The key research contribution is the combined study of the impact of data on both forecast accuracy and the resulting operational performance of MPC. This is important as it allows energy system designers to assess the trade-off between the cost of data for forecasting and the operational benefits it provides. The impact of aspects of data on forecast accuracy not yet studied in the context of building energy management are also investigated, specifically the reuse of prediction models, selection of model data features, and online model training. Two strategies for improving the efficacy of collected data for building load forecasting are proposed: a load profile similarity metric for selecting prediction models for reuse, and a change-point detection based methodology for screening training data to improve prediction performance whilst reducing model training time. Further, a long duration (10 year) historic building energy dataset is used to conduct these experiments, allowing performance to be evaluated on a multi-year scenario, providing more robust testing than existing studies.

The remainder of this chapter is structured as follows. Section 4.2 describes the simulation framework and data used to perform the experiments, as well as the models tested, and how their performance is evaluated. In Section 4.3.1, the forecast accuracy of the prediction models trained without data duration limitations is evaluated to provide a baseline, and model performance is compared. Model generalisation for load prediction between buildings is then tested in Section 4.3.2, to assess whether model reuse is a viable strategy for reducing data collection requirements for new smart energy storage systems. A load profile similarity metric based on the Wasserstein distance between functional Principal Component Analysis (fPCA) coefficient distributions is proposed, and its efficacy as a criterion for selecting models for reuse is tested. Section 4.3.3 studies the impact of various aspects of data on forecasting accuracy. The effect of the volume of data used for model training is studied in Section 4.3.3 to support decision making on the quantity of data that should be collected for model development. A change-point detection based methodology for screening training data is proposed, and its ability to improve prediction accuracy whilst reducing training data durations is investigated. The selection of model data features and use of online model training are considered in Sections 4.3.3 and 4.3.3 respectively. Section 4.3.4 contextualises the study of model forecast accuracy in the building energy system control task by quantifying the relationship between forecast accuracy and the resulting MPC operational performance for synthetic noisy forecasts. Finally, conclusions are drawn in Section 4.4.

4.2 Experimental setup

4.2.1 Smart building control simulation framework

To study the impacts of data on forecasting and MPC performance in the context of building energy systems, a case study of an example multi-building building energy system was simulated. In this case study, MPC is used to schedule battery storage operation in the multi-building energy system containing distributed storage and solar generation, as to reduce the electricity price, carbon emissions, and grid impact associated with meeting the electrical demand of the buildings. The CityLearn (Vazquez-Canteli, Dey, et al., 2020; Vázquez-Canteli et al., 2019) building energy control framework is used to simulate the behaviour of the building energy system, and provide the required data to a Linear Program based MPC implementation. A schematic of the energy flows within the simulated multi-building energy system is provided in Fig. 4.3.

During simulations, at each time step, the prediction models use observation data to produce forecasts of the operational variables, which are passed to a linear predictive control model. The resulting linear optimisation problem is solved to determine the optimal control actions, which are then applied to the battery units in the CityLearn simulation. The combination of prediction models and linear predictive control model comprise the Linear MPC controller. This simulation and control loop is illustrated in Fig. 4.2.

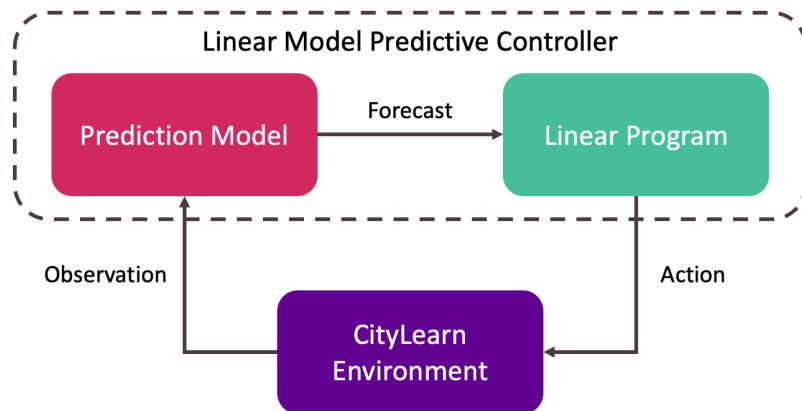


Fig. 4.2 Control loop for test district-level building energy system simulation.

The Linear Program formulation used in the MPC scheme is described by Eq. 4.1, with Table 4.1 providing descriptions of the parameters. At each time step the optimised control actions, $E_i^*[\tau=0]$, are taken. The optimisation objective is comprised of three weighted components, which correspond to the cost of grid electricity consumed by the buildings (assuming no net electricity metering), the embodied carbon emissions associated with the grid electricity, and the ramping of the overall grid electrical demand which represents the grid impact. The three components are normalised by the values they would have if no battery storage were present in the buildings, denoted by \tilde{O}^k for component k , lower bounded at 1. This clipping is performed to prevent ill-conditioning of the objective when the no-storage objective values are small.

The CityLearn simulations are configured so that the building energy system has linear dynamics, which is possible as a solar-battery system is studied and only electrical behaviour is considered. As the system parameters are known to the controller, the MPC has a perfect model of the true system dynamics. This perfect match between the simulator and controller dynamics models means no inaccuracies are introduced in system identification. The optimality guarantees of Linear Programming, and the use of the global operational objective as the MPC objective, with sufficiently large planning horizon T , mean there is negligible distortion of the operational performance from these factors. This allows the effect of operational condition forecast accuracy on MPC operational performance to be studied in isolation - i.e. to a good approximation, sub-optimality in operational performance in the simulation environment is caused solely by forecasting inaccuracies.

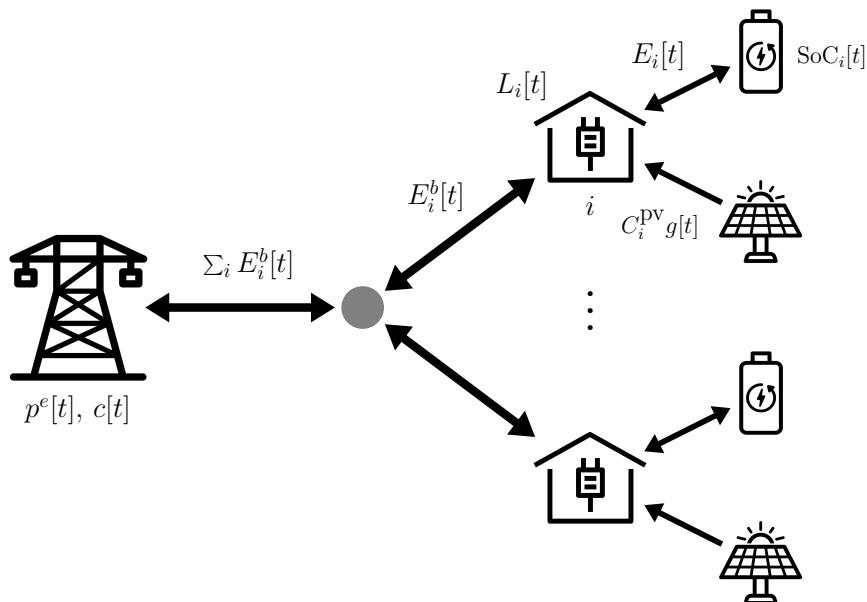


Fig. 4.3 Energy flow schematic for test multi-building energy system. (Icon credits: [Symbolon](#))

$$\begin{aligned}
\min \quad & \frac{\gamma^p \sum_{\tau} p^e[t+\tau] \sum_i (\max [0, E_i^b[t+\tau]])}{\max [1, \tilde{O}^p]} \\
& + \frac{\gamma^c \sum_{\tau} c[t+\tau] \sum_i (\max [0, E_i^b[t+\tau]])}{\max [1, \tilde{O}^c]} \tag{4.1}
\end{aligned}$$

$$+ \frac{\gamma^r \sum_{\tau} \left| \left(\sum_i E_i^b[t+\tau] \right) - \left(\sum_i E_i^b[t+\tau-1] \right) \right|}{\max [1, \tilde{O}^r]}$$

over $E_i[\tau], \text{SoC}_i[\tau+1] \quad \forall i, \tau$

subject to $\text{SoC}_i[\tau+1] \leq \text{SoC}_i[\tau] + \min [E_i[\tau]\sqrt{\eta_i}, E_i[\tau]/\sqrt{\eta_i}] \tag{4.1a}$

$$- P_i^{\max} \Delta t \leq E_i[\tau] \leq P_i^{\max} \Delta t \tag{4.1b}$$

$$0 \leq \text{SoC}_i[\tau+1] \leq C_i^s \tag{4.1c}$$

$$\text{SoC}_i[\tau=0] = \text{SoC}_i^t \tag{4.1d}$$

$$E_i^b[\tau] = L_i[\tau] - C_i^{\text{pv}} g[\tau] + E_i[\tau] \tag{4.1e}$$

for all $i \in [0, B-1], \tau \in [0, T-1]$

Parameter	Units	Description
<i>Decision variables</i>		
$E_i[\tau]$	kWh	Energy <i>intake</i> to battery unit in building i at time τ in planning horizon
$\text{SoC}_i[\tau]$	kWh	State-of-charge of battery unit in building i at time τ in planning horizon
<i>Data variables</i>		
Δt	hrs	Time step of simulation data
C_i^s	kWh	Energy capacity of battery unit in building i
C_i^{pv}	kWp	Peak power capacity of solar PV unit in building i
η_i	–	Round-trip efficiency of battery unit in building i
P_i^{\max}	kW	Power capacity of battery unit in building i
SoC_i^t	kWh	State-of-charge of battery unit in building i at time t
$L_i[t]$	kWh	Electrical demand of building i at time t
$g[t]$	kW/kWp	Normalised generation power from solar PV at time t
$p^e[t]$	£/kWh	Grid electricity price at time t
$c[t]$	kgCO ₂ /kWh	Carbon intensity of grid electricity at time t
γ^p	–	Fraction of objective from electricity cost component
γ^c	–	Fraction of objective from CO ₂ emissions component
γ^r	–	Fraction of objective from grid ramping component

Table 4.1 Description of Linear Program model parameters.

4.2.2 Forecasting task & performance evaluation

For the MPC scheme used, Eq. 4.1, forecasts of the following operational condition variables over the planning horizon T are required at each time instance, t :

- electrical demand for each building, $L_i[t]$
- normalised solar PV generation power, $g[t]$
- price of grid electricity, $p^e[t]$
- carbon intensity of grid electricity, $c[t]$

This study investigates the impacts of data on the forecasting of these 4 types of operational variables. The accuracies of these forecasts are quantified using two error metrics, the normalised Mean Absolute Error (nMAE), and the normalised Root Mean Squared Error (nRMSE), given by Eqns 4.2 & 4.3,

$$\frac{\frac{1}{N} \sum_{t=0}^{N-1} \frac{1}{T} \sum_{\tau=1}^T |f_{t,\tau}^v - v_{t+\tau}|}{\frac{1}{N} \sum_{t=0}^{N-1} v_t} \quad (4.2)$$

$$\frac{\frac{1}{N} \sum_{t=0}^{N-1} \sqrt{\frac{1}{T} \sum_{\tau=1}^T (f_{t,\tau}^v - v_{t+\tau})^2}}{\frac{1}{N} \sum_{t=0}^{N-1} v_t} \quad (4.3)$$

where $f_{t,\tau}^v$ is the forecast of variable v at time t for time instance τ in the planning horizon, and $v_{t+\tau}$ is the true value of the target variable at time instance $t + \tau$. These error metrics are the means of the standard MAE and RMSE errors over all forecasting horizons considered in the simulation, normalised by the mean level of the target variables to allow comparability between forecast accuracies.

The operational performance achieved by the MPC scheme using a given set of prediction models is quantified by evaluating the objective specified in Eq. 4.1 with the simulated behaviour of the building energy system resulting from the use of the controller.

All experiments conducted in this study test the same multi-building energy system, with distributed energy assets as specified in Appendix 4.A, and use a planning horizon of $T = 48\text{hrs}$, justification of this value is provided in Appendix 4.B, along with objective component weights, $(\gamma^p, \gamma^c, \gamma^r) = (0.45, 0.45, 0.1)$, in both the MPC scheme and for operational performance evaluation.

4.2.3 Building electricity usage data

Historic building electricity metering data from the Cambridge University Estates dataset ([Langtry and Choudhary, 2024](#)), described in Appendix A, is used to provide a case study for the multi-building energy system. A 10 year period from 2010 to 2019 is chosen for the case study, so that data durations up to a maximum practical length can be tested. 15 buildings⁴³ with complete data availability for electrical load over this period are selected⁴⁴ for use in the experiments, such that they cover a wide range of building scales and provide a good mix of similarity and dissimilarity with respect to their electricity loads.

Alongside the building electrical load data, which is assumed not to contain any significant contributions from space heating or cooling, the following other data variables are retrieved from the dataset: weather data for Cambridge (including outdoor temperature, relative humidity, and direct & diffuse solar irradiance), local solar generation power, grid electricity price & carbon intensity data, and temporal information (including hour, day, month indices, and daylight savings status).

The 10 years of data is initially split into training, validation, and test datasets covering the following periods: train (2010 to 2015), validate (2016 to 2017), test (2018 to 2019). For all experiments the test data is kept the same, however the periods of data used to train the prediction models are altered in Section 4.3.3.

4.2.4 Brief description of the prediction models

Data and computational requirements, which vary across predictions models, are important considerations for the deployment of MPC based controllers in practical building energy systems. This work investigates the performance of 6 machine learning based prediction models which span a range of model characteristics; 3 simple neural models, and 3 high-complexity, state-of-the-art models. A brief description of each model architecture follows. Technical specifications of the model implementations used are provided in Appendix 4.C.

Simple neural models

Recent literature, ([Zeng et al., 2023](#)), has shown that simple neural models using Direct Multi-Step forecasting (DMS), where all predictions over the forecast window are generated concurrently, can outperform complex, transformer-based models using traditional, Iterated Multi-Step forecasting (IMS), in which a single-step forecaster is applied iteratively to generate a multi-step forecast. For all three simple neural architectures investigated, DMS forecasting is used.

⁴³ The number of buildings that can be used in the study is limited by the computational cost of the forecasting methods and the system simulations.

⁴⁴ Version 1 building numbers: 0, 3, 9, 11, 12, 15, 16, 25, 26, 32, 38, 44, 45, 48, 49

Linear neural network (Linear). A Multi-Layer Perceptron (MLP) model that maps the inputs directly to the output without an activation function (non-linearity).

Residual multi-layer perceptron (ResMLP). A Residual MLPSkip model (MLP model with skip-connections), comprised of a single hidden layer with 128 neurons.

Convolutional neural network (Conv). A Convolutional Neural Network (CNN) model that contains convolution layers followed by a linear layer. The architecture used comprises two layers with kernel sizes of 6 and 12, with five and one channels, respectively.

State-of-the-art machine learning models

Temporal Fusion Transformer. The Temporal Fusion Transformer (TFT) model (Lim et al., 2021), developed by Google, is an attention-based architecture that enables the fusion of data from multiple input sources to inform predictions. The neural structure contains features which allow for the learning of multiple underlying relationships across temporal scales, and the attention mechanism allows for interpretation of the model predictions, i.e. which data the model is exploiting to produce its forecasts. The model uses categorical covariates of date-time information, as well as temperature information for predicting building loads.

Neural Hierarchical Interpolation for Time Series Forecasting. Neural Hierarchical Interpolation for Time Series Forecasting (NHiTS) (Challu et al., 2023) is an MLP model which learns a set of basis functions at different frequencies that describe the underlying patterns in the training data, and produces forecasts by using hierarchical interpolation to combine predictions from the basis functions in a computationally efficient manner. It uses categorical covariates of date-time information.

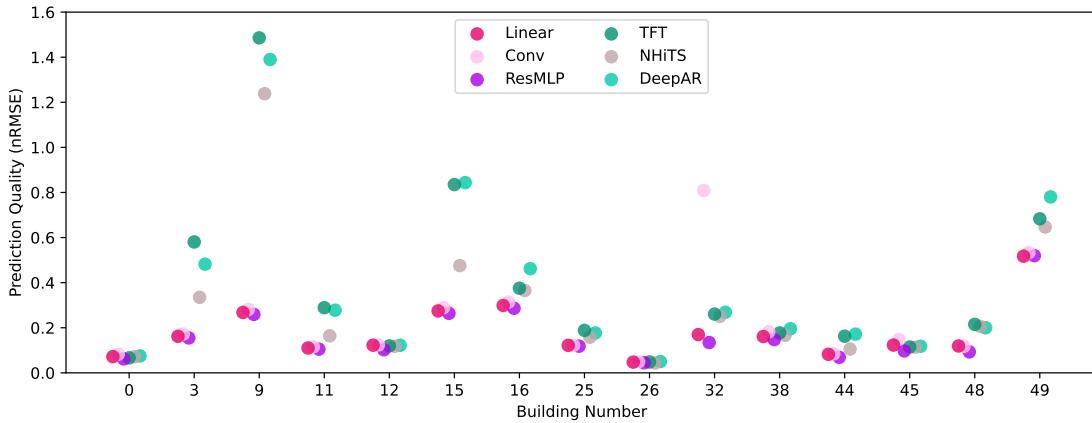
DeepAR. DeepAR is a Recurrent Neural Network (RNN) based model developed by Amazon (Salinas et al., 2020), which has been widely applied in a range of research areas. It is a probabilistic forecasting model, but for this study only the mean prediction is used. The model uses categorical covariates of date-time information.

4.3 Results & Discussion

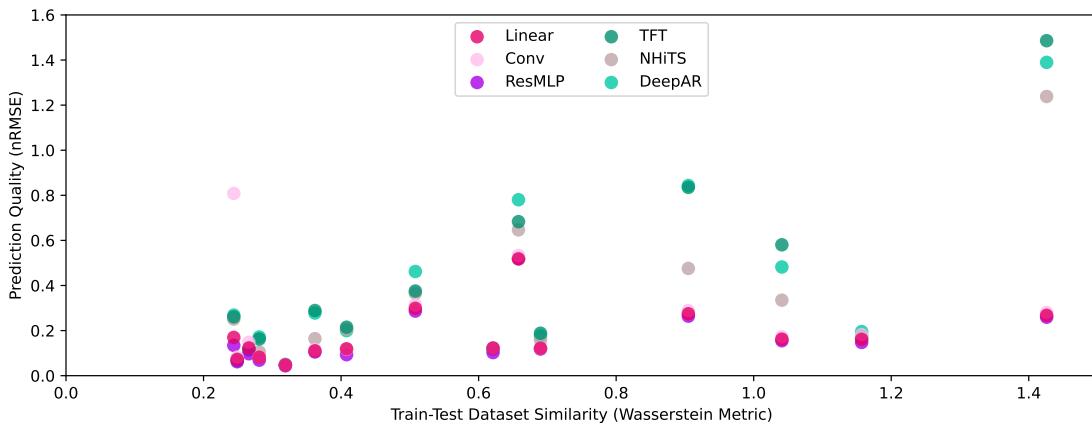
4.3.1 Baseline prediction accuracy comparison

The prediction accuracy of each forecasting model trained using 8 years of training data, the maximum available, was evaluated to provide a baseline for the model architectures in a setting without data limitations. For brevity, forecast accuracy results are discussed for the nRMSE metric only, however equivalent results were found with the nMAE metric. Figures 4.4a & 4.4c show that the simple Linear model achieves similar or better prediction accuracy (lower nRMSE values) compared to the complex, state-of-the-art models across all prediction variables. The Conv and ResMLP models provide similar accuracy when predicting building electrical loads, however both have significantly worse accuracy for the electricity price and carbon intensity prediction variables. Complex models achieve slightly better accuracy for electricity price and solar generation predictions, at most 8.5% and 4.2% lower nRMSE than the Linear model, achieved by NHiTS and TFT respectively. However, for some buildings the complex models exhibit very poor forecast accuracy for load predictions. These instances of poor accuracy are found to correlate with a measure of the similarity between the train and test datasets for building load, called the ‘Wasserstein similarity metric’, which is described in Appendix 4.D and is used for the study of model reuse in the following section. Fig. 4.4b plots the relationship between prediction accuracy and data similarity for each model, and shows that complex models provide poor prediction accuracy when the training and test data are significantly different. Hence, simple neural models provide better prediction generalisation under changes in building load dynamics between the training and test data, which is highly advantageous for application to practical systems, as occupant driven load dynamics may change after system installation, e.g. due to a change of building use.

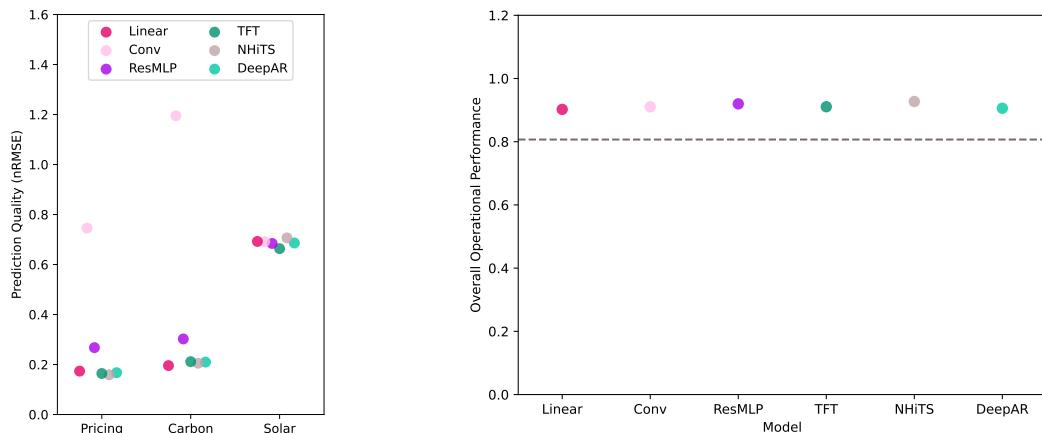
These results indicate that simple neural models have sufficient expressivity to well capture the underlying trends in building energy usage, but learn sufficiently simple relationships about the data as to avoid the problem of over-fitting experienced by the complex models, leading to better generalisation in time. This suggests that the majority of the trends in the load data are present within the 1 week input window used by the simple models, which agrees with the strong daily and weekly trends of typical building energy usage, caused by building occupancy patterns. It is likely that annual trends are well captured by the mean value over the input window, as these trends are slow relative to the prediction length. Additionally, the predictions required by MPC are relatively short in length, 48 hours in this case, whereas the complex models are found to provide more



(a) Comparison of baseline load prediction accuracy of forecasting models.



(b) Correlation between model prediction accuracy and train-test data similarity metric value (lower Wasserstein metric means more similar data).



(c) Pricing, carbon, solar prediction comparison.

(d) Comparison of MPC operational performance using baseline forecasting models.

Fig. 4.4 Comparison of baseline prediction accuracy and operational performance of forecasting models.

stable predictions and greater overall accuracy for longer duration forecasts (Challu et al., 2023; Hao Zhang et al., 2022).

Overall, simple neural models are able to provide analogous prediction accuracy to complex models across all prediction variables, and do so with substantially lower computational requirements. Table 4.2 provides the training time and inference time (time to generate forecasts) of the baseline models, and shows that simple models require roughly 500x less computation for inference. Hence, in practical systems the use of simple models would enable shorter prediction intervals, leading to higher frequency control, and allow for the use of lower cost compute hardware. Further, due to their high computational cost, the complex models were only able to use a 72 hour input window (López Santos et al., 2022). As a result, they have less information available to inform their predictions, likely limiting the accuracy they could achieve. Therefore, the computational efficiency of simple neural models enabling the use of longer input windows provides an additional advantage.

Model	Linear	Conv	ResMLP	TFT	NHiTS	DeepAR
Training time (hrs)	12.3	26.8	13.4	8.9	15.2	9.3
Prediction inference time (s)	29	73	146	15,222	15,289	74,991

Table 4.2 Computation times for baseline models, trained on 8 years of data, and predicting for simulations of 2 years duration.

The similar prediction accuracies of the tested models are found to lead to similar operational performance when used in the MPC scheme, see Fig. 4.4d, where the dashed line indicates the bound on operational performance achieved by an MPC scheme with perfect forecasts. The use of the MPC controllers with the specified battery systems leads to an average 8.7% improvement in operational performance for the multi-building energy system, with a range of 7.3% to 9.8%.

4.3.2 Model generalisation

When a solar-battery system using MPC is installed, high-resolution historic load metering data (e.g. from smart meters) may not be available for the building. In this case, the project must either be delayed to allow time for data collection, incurring a significant cost, or a prediction model trained on load data from another building must be used, potentially incurring an operational performance penalty due to worse prediction accuracy. Model reuse greatly reduces data collection requirements and the associated costs, however its appropriacy depends on the ability of load prediction models to generalise between

buildings, and the trade-off between data cost savings and increased operational cost due to lower forecast accuracies.

The generalisation of prediction models between buildings is tested by using the baseline models trained on data from each of the 15 buildings to forecast electrical load for every other building. Fig. 4.5 shows the distribution of relative forecast accuracies achieved by each model over all buildings (DeepAR is excluded due to excessive computational costs), where the y-axis plots the nRMSE forecast accuracy of the tested model (potentially trained on a different building) normalised by the forecast accuracy achieved by the model trained on data from the target building.

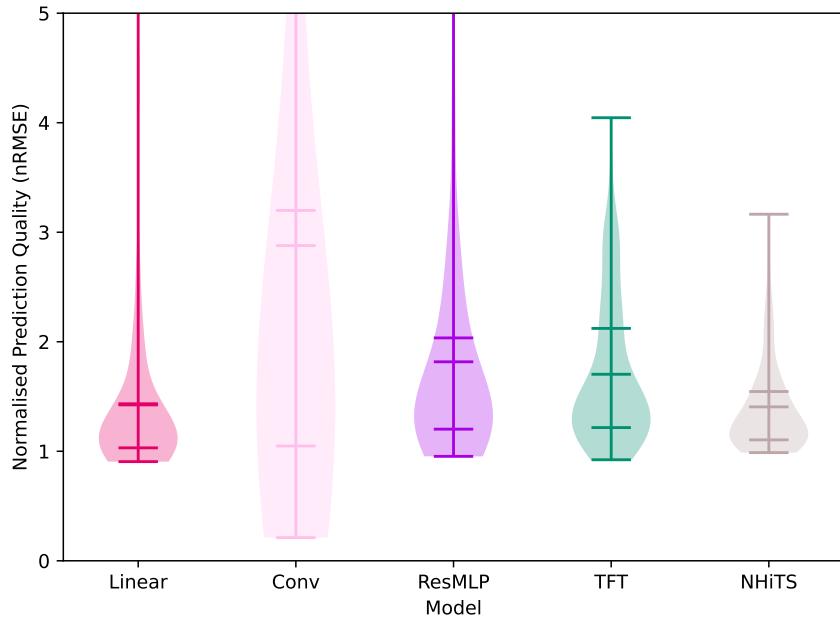


Fig. 4.5 Comparison of forecasting model generalisation. Violin plot with quartiles indicated by horizontal lines.

The Linear model provides similar generalisation performance to NHiTS, and is substantially better than all other models. It is proposed that the Linear model, being mechanistically equivalent to linear regression, achieves good generalisation as it learns relatively simple relationships about the load data, which are consistent between buildings. Whereas NHiTS achieves good generalisation due to its behaviour of generating smooth forecasts (Challu et al., 2023), making it less susceptible to producing erratic and inaccurate predictions with unseen input data. As before, TFT suffers from over-fitting, leading to relatively poor generalisation. Whilst the Conv model was able to provide good generalisation in time where the dataset similarity was close, indicated by small Wasserstein metric values in Fig. 4.4b, between buildings the load data is much less similar, and the relationships learned by the Conv

model are no longer valid, leading to poor prediction accuracy. It is suggested this is due to the features learned by the pooling process no longer being pertinent for the new building.

If a pre-trained Linear model is selected for reuse at random, the average prediction accuracy is substantially worse than a model trained on data collected from the target building (45% higher nRMSE, corresponding to the mean of the violin plot). However, in a real scenario where an ensemble of pre-trained models is available, the smart energy storage system designer could reuse a model from a building that is expected to be similar to the target building (e.g. by size, usage type, envelope characteristics, load dynamics, etc.), to maximise the likelihood of good generalisation. A load profile similarity metric is additionally proposed as a way off assessing the similarity of building load dynamics for the purposes of model reuse. This similarity metric is based on the similarity of the distributions of underlying mode shapes within each load profile, and is referred to as the 'Wasserstein similarity metric' as it uses the Wasserstein distance between mode shape coefficient distributions. Appendix 4.D describes the calculation of similarity metric values. The correlation between generalisation of the Linear model and the Wasserstein similarity metric values between the training datasets of the reused model building and target building is shown in Fig. 4.6. The cluster of points in the bottom left indicates that buildings with similar load profiles (low Wasserstein metric values) can provide models that achieve good generalisation. When selecting models for reuse by minimising the Wasserstein metric, the average relative prediction accuracy improves to an 11% increase in nRMSE, with a range of -7.9% to +52.6%, showing that some reused models outperform those trained on the true building training dataset.

These results show that when developing a load forecasting model for a new building system where little or no data is available, the reuse of forecasting models trained on existing buildings provides a promising⁴⁵ alternative to the collection of data and training of a building specific model, and the associated costs. However, reusing prediction models increases uncertainty in the operational performance the MPC system will achieve, as models trained on load data from buildings similar to the target building do not always provide good prediction accuracy, as shown by the wide range of reused model accuracies.

⁴⁵ Further work is required to determine the feasibility of identifying similar buildings from the model ensemble in practice. Whilst a small sample of load data from the target building could be used to approximate the Wasserstein similarity metric, the quality of this approximation must be weighed against the forecast accuracy that could be achieved by a model trained using that data. Though, additional information on the building, such as usage type, could be used to guide model selection for reuse.

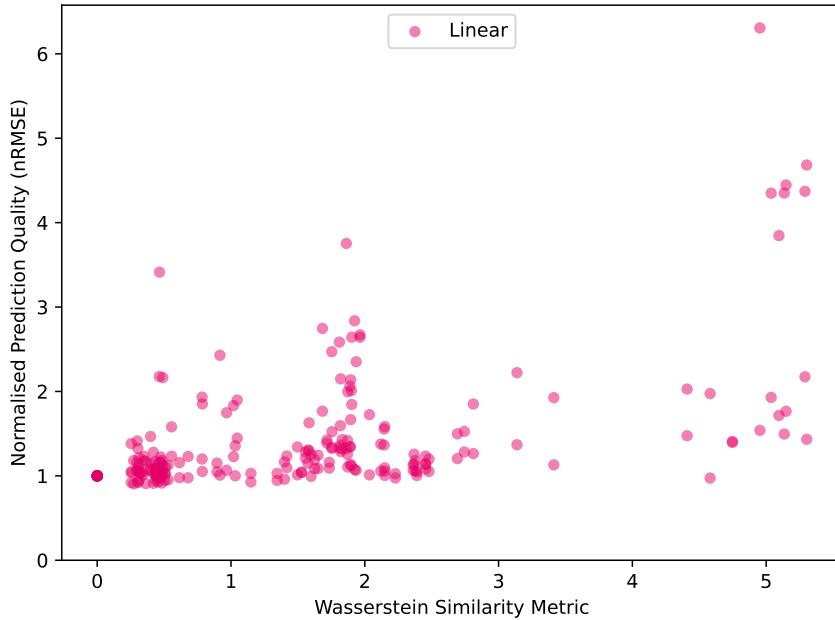


Fig. 4.6 Correlation between Linear model generalisation and training dataset similarity metric values.

4.3.3 Data efficiency

Data required for developing forecasting models incurs cost from both its acquisition and the computation required to exploit it. However, using longer durations of training data or additional data variables can increase forecast accuracy, resulting in lower operational cost of the energy system. Therefore, any measures to reduce the quantity of data used must be traded-off against the corresponding decrease in prediction accuracy. This section explores a series of aspects of data efficiency and their effect on forecast accuracy. Subsequently, Section 4.3.4 investigates the impact of forecast accuracy on operational performance of the MPC.

Testing data efficiency requires several versions of a prediction model to be trained. Due to computational limitations, for some data efficiency tests only the Linear prediction model is studied, as the complex models are too computationally expensive, and results from previous sections show the Linear model is the best performing simple model.

Volume of training data

Using greater volumes of training data, longer durations of historic measurements of buildings, can improve model prediction accuracy by avoiding over-fitting, improving temporal generalisability, but comes at a roughly linear cost. The trade-off between the length of training data used, the combined duration of the train and validate datasets, and forecast

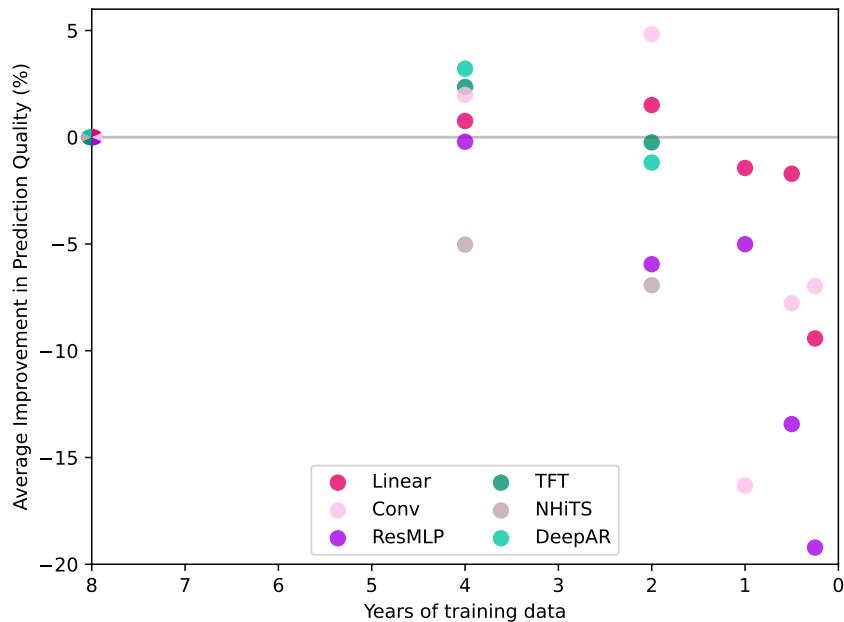


Fig. 4.7 Average improvement in prediction accuracy over all prediction variables with years of data used for training.

accuracy is investigated by training prediction models on subsets of the building energy dataset of varying durations from 8 years, the maximum duration available, to 3 months (one season), the shortest duration tested in (Choi and S. Lee, 2023). For the complex models, at least 2 years of training data was required due to the use of temporal covariates. Fig. 4.7 plots the proportional improvement in prediction accuracy of each model (negative improvement shows worsening model accuracy) compared to baseline when training using different data volumes. Across a broad range of model architectures, reducing training data length down to 2 years has a limited impact on prediction accuracy, in some cases improving prediction accuracy. Trends differ between prediction variables, however for the Linear model, in the majority of cases there is a small prediction accuracy penalty between 8 and 1 years of training data, and then a rapid worsening of forecasting accuracy below 6 months. This indicates that when making data collection decisions to support prediction model development for building electrical energy systems, at most 2 years of measurement data should be gathered. However, at least two seasons of data (6 months) are required to prevent model over-fitting and learn trends which generalise across seasons.

Screening training data using change-points. Once building data has been gathered, there remains a question as to whether all of the available data should be used for training, or whether some data should be excluded to improve model prediction. For instance, if data that is non-representative of the present building load dynamics can be excluded,



Fig. 4.8 Improvement in prediction accuracy of models trained using data from change-points onward.

prediction accuracy can be improved alongside data efficiency. Change-point analysis can be used to screen training data for this purpose, by detecting changes in building load dynamics and excluding data preceding the change from training. A change-point analysis using the BEAST algorithm (Zhao et al., 2019) was performed on the building load dataset, described in Appendix 4.E. Points where changes in load trend were detected were used to screen the training data for 7 selected buildings. In order to allow a better analysis of the detected change-points, the training dataset was increased to 7 years (2010-2016), leaving 1 year of validation data (2017). Linear prediction models were trained using sections of the training data starting at each detected change-point and running to the end of 2016.

Fig. 4.8 shows the relative improvements in prediction accuracy of models trained on data screened using change-points, compared to the case where the full 7 years of available data is used, where the x-axis position indicates the timing of the change-point. For most buildings, screening the training data using change-points improves prediction accuracy, indicating that non-representative data is being removed from the training dataset. This suggests that ex-post training data screening can be used to improve model accuracy whilst reducing the quantity of training data used, and so computational cost. These results correspond with the findings of (Choi and S. Lee, 2023), which shows that additional training data only improves prediction accuracy if it is sufficiently similar to the test data, and can reduce accuracy when insufficiently similar. However, the results of this study demonstrate an additional consideration, which is that prediction accuracy

worsens substantially when insufficient training data volume is used. This suggests a trade-off between having enough data to avoid over-fitting, and removing non-representative data which causes over-generalisation of the model. Figures 4.7 & 4.8 indicate that at least 1 year of training data is required to achieve good prediction accuracy, after which only sufficiently similar data should be included in the training dataset. The reason for the difference in results compared to (Choi and S. Lee, 2023) is likely due to this study testing model prediction accuracy on 2 years of data, compared to 0.1 years in (Choi and S. Lee, 2023). As a result, models in this study must capture the seasonality of building load dynamics, hence requiring at least 1 year of training data to achieve good prediction accuracy.

The data screening method considered in this study is relatively simple, using change-points to exclude data by classifying it as ‘non-representative’ of current behaviour. More advanced techniques could be created by quantifying the similarity between the sections of data identified using the change-points and the validation data, e.g. using the Wasserstein similarity metric, and selecting sections of data to use in the training dataset using both the probability of a true change having occurred and the data similarity metrics.

Change-point analysis can also be used online to detect real time changes in building load dynamics and provide indication of when the prediction models should be updated (i.e. retrained using recent data), due to the training data of the current model no longer being representative of the building behaviour, causing reduced prediction accuracy. This is particularly pertinent in the context of climate change, which is expected to significantly impact the energy usage behaviour of buildings, for instance reducing peak heating loads due to higher winter temperatures, and increasing summer electrical loads to provide cooling during prolonged heat waves.

Data features

The variables that can be monitored in building energy systems are often well correlated, for example ambient temperature and electrical load. As a result, covariates can be used by prediction models to attempt to identify underlying links between the variables and improve forecast accuracy. However, the use of additional data variables (features) incurs a cost from both the collection and exploitation of the extra data, for instance the purchase of proprietary weather forecasts. The impact of feature selection on forecast accuracy is investigated by comparing the prediction accuracy of models trained using varying numbers of data features. The Pearson correlation coefficient between the 11 data features available in the building energy dataset are shown for the case of Building 0 in Fig. 4.9. Linear prediction models were trained using the n most correlated variables/features for each of the electrical load, solar generation, electricity price, and carbon intensity target

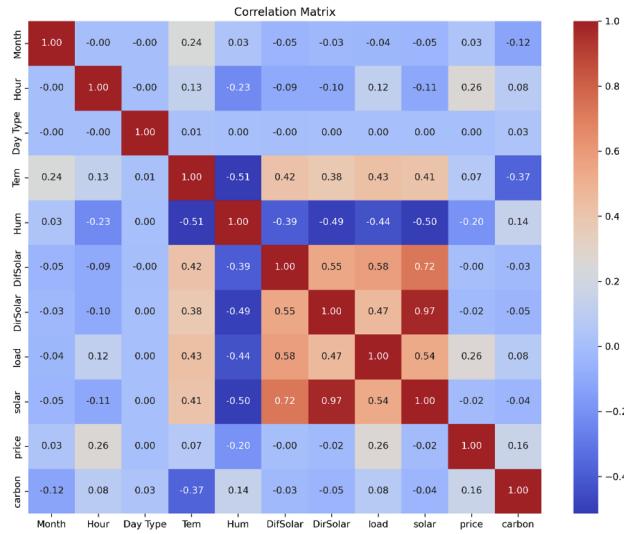


Fig. 4.9 Pearson correlation between data variables for Building 0.

variables. Feature selection was performed separately for each case by ranking Pearson correlation coefficients between the target variable and all other available covariates. Fig. 4.10 shows the impact of the number of data features used by the models on forecast accuracy. For most prediction variables, the inclusion of additional data features worsens forecast accuracy. Several factors could contribute to this behaviour:

- Over-fitting (Hawkins, 2004) : The incorporation of an excessive number of features may lead the model to over-fit the training data, worsening its predictions for unseen data.
- Multi-collinearity (Farrar and Glauber, 1967) : The presence of highly correlated features can destabilize the model, resulting in worse prediction accuracy.
- Curse of Dimensionality (Verleysen and François, 2005) : An increase in the number of features increases the dimensionality of the data, which can lead to data sparsity and degradation of prediction accuracy.

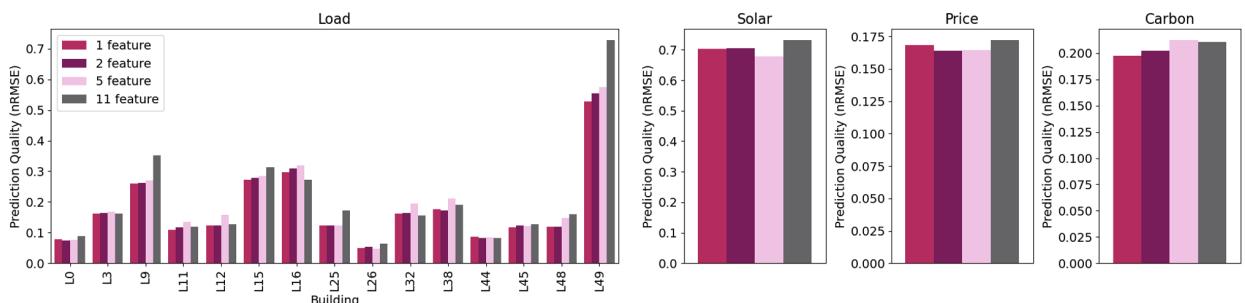


Fig. 4.10 Variation of prediction accuracy with no. of data features included in Linear model.

These results show that testing should be performed before decisions are made regarding the collection of covariate data to ensure that, for the type of system in question, its use will firstly improve prediction accuracy, and secondly that said accuracy improvements warrant the cost of the data. This study considers a simple, correlation based feature selection method, however more advanced, optimization based techniques (González-Vidal et al., 2019; W. Kim et al., 2020) could be used to determine the set of available features which provides the optimal trade-off between data cost and model prediction accuracy.

Online training

During building operation monitoring systems continuously collect operational data. The characteristics of building behaviour can change during operation due to external factors such as weather & climate, occupancy, and equipment degradation & maintenance. Continuous online training updates the predictive models using the collected monitoring data, improving the model's ability to adapt to dynamic changes in building behaviour. A version of the Linear prediction model with online training is implemented to investigate the impact of online training frequency on model accuracy. The model is updated online by tuning the model parameters (retraining) using a sliding window of training data with length equal to the update frequency. The prediction accuracy of models updated at different frequencies is plotted in Fig. 4.11. Higher update frequencies led to increased prediction accuracy across all prediction variables. For example, prediction accuracy for grid electricity price improved by 15.7%, 14.0%, 12.1%, and 8.1% when updated monthly, quarterly, semi-annually, and annually, respectively, compared to the baseline model which is not trained online. Online training allows the prediction model to adapt to changes in the underlying trends which occur over time, as it can learn the trends in recent data. This improves prediction accuracy as the model does not need to generalise in time, as it is trained (updated) on data that is representative of the current prediction horizon. Therefore, it is expected that the additional computational hardware and system complexity required for online trained prediction models will be worthwhile in practical systems.

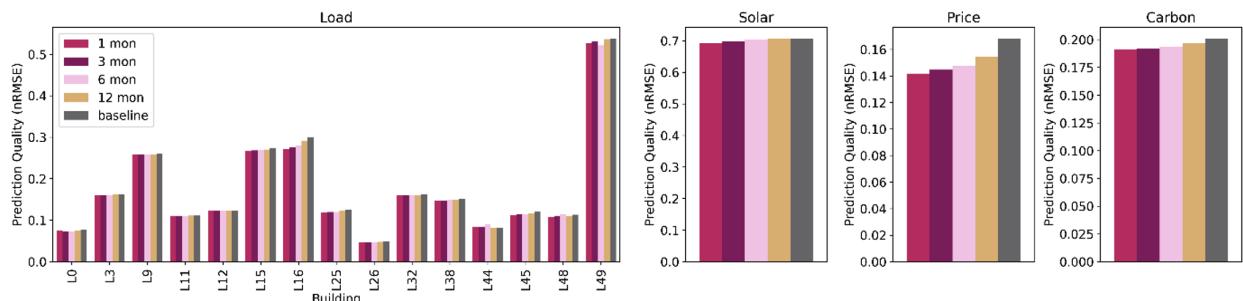


Fig. 4.11 Variation of prediction accuracy with online update frequency.

4.3.4 Sensitivity of Model Predictive Control to forecast accuracy

Whether investments in additional data to improve forecast accuracy provide net benefit to the operation of a smart energy storage system depends on the resulting improvements in operational performance achieved by the MPC controller. The determination of optimal data collection strategies therefore requires quantification of the relationship between forecast accuracy and MPC operational performance. To investigate this in a controlled setting, the operational performance achieved by the tested MPC scheme is evaluated with the use of synthetic forecasts. The synthetic forecasts are produced by adding a Gaussian random walk noise component to the ground-truth values, as described in Eq. 4.4,

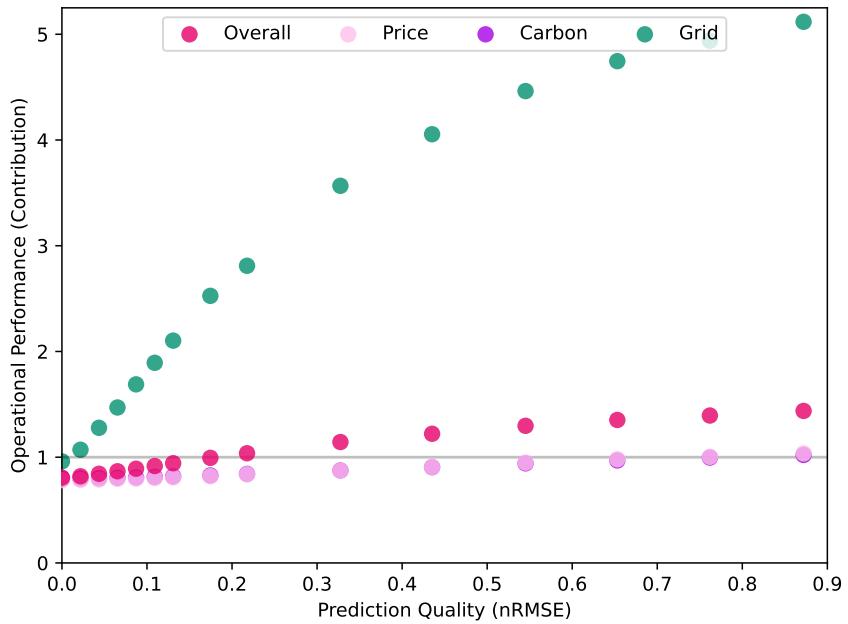
$$f_{\text{GRW}}^v[t, \tau] = v_{t+\tau} + \sum_{j=1}^{\tau} w_j^\sigma \quad \text{s.t. } w_j^\sigma \text{ i.i.d. } \mathcal{N}(\mu=0, \sigma^2) \quad (4.4)$$

where $v_{t+\tau}$ is the ground-truth value of variable v at instance τ in the planning horizon of the forecast created at time t , and the noise level σ can be selected.

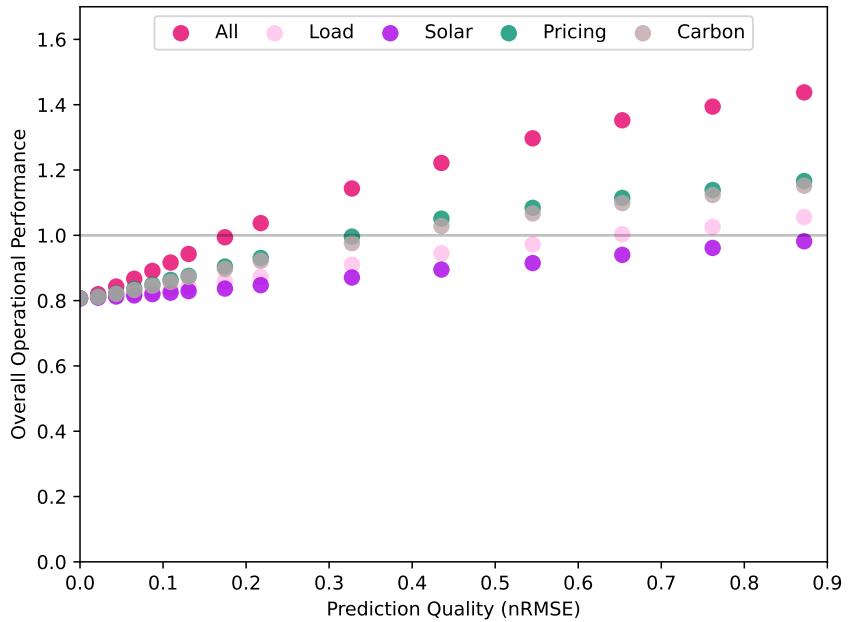
The MPC scheme using these synthetic forecasts for all prediction variables is tested at varying noise levels. Fig. 4.12a shows the resulting operational performance of MPC, as well as the three components it is comprised of (electricity price, carbon emissions, and grid impact), as the prediction accuracy of the synthetic forecasts varies. Fig. 4.12b shows the variation in overall operational performance of the MPC when synthetic forecasts are used for each type of prediction variable in turn, with perfect forecasts used for all other variables. The horizontal line at 1 indicates the performance of the building energy system without battery control, which is the point at which the MPC controller becomes redundant.

The results show that the tested MPC scheme is most sensitive to the forecast accuracy of the grid electricity price and carbon intensity variables, suggesting that the most resource and expense should be invested in producing accurate forecasts of the grid conditions. Additionally, whilst the prediction models tested in this study performed worst when forecasting solar generation, see Fig. 4.4c, it may not be worth expending additional resource to improve these forecasts as the MPC scheme is least sensitive to this variable. Though the grid component of operational performance is found to be by far the most sensitive to forecast accuracy, this is likely due to the synthetic forecast model used⁴⁶ for the development of forecasting models in practical smart energy storage systems, and is not considered to be reflective of the behaviour of real prediction models.

⁴⁶ In the synthetic forecast model used, the Gaussian random walk noise component is regenerated at each time instance. This leads to the generation of forecasts which are correct on average, but can cause the MPC controller to take opposing energy management strategies in successive time steps, leading to large magnitudes of energy flows from the battery to correct the strategy. Further work is required to create a noise model for synthetic forecasts that is calibrated to the operational performance vs forecast accuracy trade-off of practical prediction models. With such a noise model, the experimental method demonstrated could be used to quantify the economic advantage of forecast accuracy improvements to support decision making regarding data collection.



(a) Variation of operational performance components with amplitude of noise on all prediction variables.



(b) Variation of overall operational performance with amplitude of noise on each type of prediction variable.

Fig. 4.12 Sensitivity of operational performance to synthetic forecast noise simulating prediction inaccuracy.

4.4 Conclusions

This chapter investigated the impacts of data on the accuracy of machine learning based forecasting models for building operational conditions, and the resulting operational performance of a **Model Predictive Control (MPC)** scheme in a multi-building energy system with distributed generation and storage. Experiments were conducted using a large-scale dataset of electrical load measurements from buildings in the Cambridge University Estates.

A simple linear multi-layer perceptron model (Linear) using DMS prediction was found to achieve equivalent forecast accuracy to high-complexity, state-of-the-art machine learning models in a setting without data limitations, but has substantial advantages regarding data efficiency and generalisation performance. So, this simple neural model is better suited for use in practical **MPC** systems due to its lower data and computational requirements, and better performance on new load dynamics.

Using more than 2 years of hourly resolved training data did not provide significant improvements in prediction accuracy for most of the models tested, indicating that the collection of monitoring data for longer durations is unnecessary for the development of performant **MPC** schemes. Further, screening training data using change-point analysis to remove low similarity data was able to simultaneously improve data efficiency and prediction accuracy, provided at least 1 year of training data was kept. This could also be achieved through the removal of redundant data features from models.

The reuse of Linear models for load prediction between buildings was shown to be an effective way of reducing data collection requirements. When selecting models for reuse using a proposed load profile similarity metric based on the Wasserstein distance between fPCA coefficient distributions, model reuse led to an average 11% increase in prediction error. However in comparison, models trained using only 3 months of building specific data provided forecasts with an average 9.9% increase in error compared to the baseline. Additionally, online training of prediction models was shown to be highly beneficial, with higher update frequencies providing greater forecast accuracy improvements, of up to 15.7% for the case of monthly updating. Hence, monitoring data should be used to update reused prediction models *in situ* to tailor them to the building system, ultimately replacing the reused model. The results suggest that, by exploiting existing building energy datasets to pre-train models, in many cases sufficient forecast accuracy can be achieved without the collection of any building load data prior to the installation of the system.

The relationship between forecast accuracy and operational performance of the **MPC** controller was investigated for synthetic forecasts, which showed the **MPC** scheme is most sensitive to grid electricity price and carbon intensity prediction accuracy. This indicates that the most effort and expense should be spent on developing or purchasing high accuracy

forecasts of the grid conditions, as these will provide the greatest improvement in operational performance of the controller.

In this chapter, the operational performance of the MPC controller was studied using a relative performance metric⁴⁷. However, the analysis could be repeated using an economic performance metric. This would allow the financial benefit of data for improving forecast accuracy to be quantified. By comparing to the costs of data collection, building operators could determine whether expenditure on additional data collection or commercial forecasts is beneficial, i.e. whether the costs of data are outweighed by the improvements in operational performance provided. And if so, the results could also be used to determine the optimal data collection strategy.

This analysis would be conceptually quite similar to a Value of Information analysis, and has the same objective. However in this case using the Vol methodology is not appropriate. While collecting more training data could be viewed as reducing uncertainty in the behaviour of a building, it is not clear how a statistical model of this uncertainty could be constructed, and whether the setup is really compatible with a statistical viewpoint⁴⁸. The analysis in Section 4.3.3 showed that collecting more data does not always improve prediction accuracy, as building load dynamics can change over time, and additional historic data is not always representative of the current behaviour, meaning it can worsen model training. This breaks one of the key assumptions of the Vol methodology. An analysis using synthetic data could be performed, however it would be difficult to justify or validate its applicability to real-world energy systems. So due to the complexity of the data required to study forecasting accuracy, the case study based approach used in this chapter is better suited for investigating data collection requirements for forecasting.

⁴⁷ This was done to match the goal of the CityLearn Challenge 2023 (Nagy et al., 2023).

⁴⁸ Data collected from other buildings could be viewed as a prior distribution for the load data of the target building. This is the modelling approach taken in Chapter 5. A prediction model trained on this data, often referred to as a general or foundation model, could be seen as a prior decision. This general model could then be updated using data collected from the target building in a Bayesian-like fashion, improving prediction accuracy for the target building. An approach similar to this is taken in (Raisch, Langtry, et al., 2025) to study the impact of data on the accuracy of system dynamics models.

Appendix 4.A Multi-building energy system asset specification

The specifications of the distributed generation and storage assets for the multi-building energy system simulated in this study are provided in Table 4.3. The battery power capacities are set to be 3 times the mean building electrical load in the test dataset, the battery energy capacities are set to be able to provide 24 hours of that mean electrical load, and the solar PV power capacities are set by assuming 90% of the roof area of each building is filled by solar panels of power density 0.15 kWp/m².

Building	Battery power capacity (kW)	Battery energy capacity (kWh)	Battery efficiency (%)	Solar PV power capacity (kWp)
0	531	4246	90	461
3	60	478	90	38
9	14	108	90	20
11	342	2736	90	41
12	98	780	90	89
15	203	1622	90	143
16	306	2448	90	120
25	275	2198	90	118
26	1001	8006	90	136
32	157	1255	90	166
38	58	461	90	46
44	1084	8674	90	663
45	1245	9962	90	437
48	622	4973	90	500
49	57	458	90	258

Table 4.3 Specification of distributed energy assets in simulated multi-building energy system.

Appendix 4.B MPC planning horizon

The planning horizon used in a MPC scheme determines its ability to effectively arbitrage energy and meet the operational objectives of energy management. Previous studies show that a planning horizon of 24 hours is sufficient when controlling HVAC systems (Oldewurtel et al., 2012), and recommend 48 hours as a suitable value for systems with distributed solar and storage (Thieblemont et al., 2017).

To demonstrate the suitability of a planning horizon of $T = 48\text{hrs}$ for the experiments on the test system in this study, the operational performance of the MPC scheme using perfect forecasts is evaluated over varying planning horizon lengths, along with the computational time required to solve the Linear Programs for the simulations, shown in Fig. 4.13. It can be seen that $T = 48\text{hrs}$ provides near-optimal control performance at a reasonable computation time, and so is a suitable choice for the experiments.

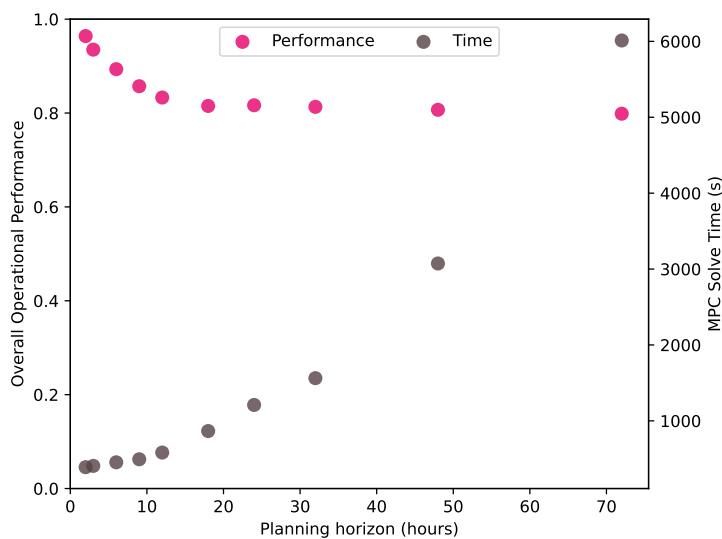


Fig. 4.13 Variation of perfect forecast MPC operational performance and computation time with planning horizon, T .

Appendix 4.C Technical specification of prediction models and training procedures

The model parameters and training procedure used for machine learning models have a significant impact on their performance, as well as their computational and data requirements. Hence, they must be selected carefully for the target application to maximise model performance with the available computational resources and data.

4.C.1 Simple neural models

For the simple neural models, implementations from the Pytorch Lighting library (Falcon and The PyTorch Lightning team, 2019) were used. The models use an input window of 168 time steps (input dimension) and output a forecast of 48 time steps (output dimension). All models were trained using the Adam optimizer (Kingma and Ba, 2017), with a batch size of 32 and a learning rate determined using the built-in learning rate tuner.

The Linear model consists of a single linear layer, of size 48, with no activation function used.

The Convolution model uses two 1D convolution layers with channels (5, 1) and kernel sizes (5, 6), respectively, with ReLU activation functions. This is followed by an output MLP that brings the dimensionality to 48.

The ResMLP model has an linear layer with 168 nodes, using a residual connection and ReLU activation. A subsequent output linear layer brings the dimensionality to 48.

4.C.2 State-of-the-art machine learning models

For the complex, state-of-the-art models, implementations from the Pytorch Forecasting library (Beitner, 2024) were used. The models take in data from the previous 72 time steps along with known covariate data from the coming 48 time steps, and output forecasts of the target variable for the coming 48 time steps. All models were trained using an early stopping trainer with a patience of 5, for a maximum of 50 epochs, with a limit of 800 training batches per epoch. The learning rate was determined using the built-in learning rate tuner. The size parameters, loss functions, and optimizers used are detailed in Table 4.4.

Model	TFT	NHiTS	DeepAR
Hidden size	48	128	64
Drop-out	0.1	0.1	0.1
Loss	Quantile loss	Quantile loss	Normal distribution loss
Optimizer	Adam	AdamW	Adam
	Attention head size: 4	Weight decay: 0.01 Backcast loss ratio: 0	RNN layers: 3

Table 4.4 Technical specifications of complex, state-of-the-art models.

Appendix 4.D Data similarity metric

4.D.1 Functional data analysis

Functional Data Analysis (FDA) is used to analyse and compare the energy usage behaviours across the building electrical load dataset of 15 buildings. In this approach, functional principal components (fPCs) are extracted from the data, such that each data sample can be constructed from an equation of the form,

$$f(t) = \mu(t) + \sum_{i=1}^n \alpha_i \nu_i(t) \quad (4.5)$$

i.e. the data sample, $f(t)$, is constructed from a linear sum of a mean function, $\mu(t)$, and a weighted sum of n fPCs, $\nu_i(t)$, with weightings or ‘scores’ α_i . The mean function and fPCs are the same across all data samples, whereas the weightings are unique to each data sample. This has the benefit that to compare data samples, which are functions of time, it suffices to compare the weightings, a low-dimensional representation of the functional data, that can be analysed using standard statistical techniques.

For this analysis, the datasets are pre-processed into daily time histories, such that each data sample is a 24 hour profile of electricity consumption, starting at midnight. For each building, the training dataset comprises 6 years of data and hence 2191 data samples, and the validation and test datasets are both 2 years and hence 731 and 730 data samples respectively (2016 was a leap year).

The functional Principal Component Analysis (fPCA) approach used involves first aligning the data samples to a common mean, $\mu(t)$. This generates a warping function and an amplitude function for each data sample that describe how the data sample maps to the

mean function. The warping function describes the phase relationship, i.e. the variation in time, and the amplitude function describes changes in magnitude. The warping and amplitude functions are then analysed separately and fPCs generated for both. The approach is illustrated schematically in Fig. 4.14, and full details of the approach are described in (Ward, 2021) (Ward, 2021).

Fig. 4.15 illustrates the first two phase and amplitude fPCs extracted from the dataset. Fig. 4.16 shows distributions of example fPC coefficients for Buildings 38 and 49. The load data for Building 49 exhibits a much lower range than that for Building 38, which results in a more positive distribution of V1 fPC coefficients. In Fig. 4.15 positive coefficients V1 fPC can be seen to reduce the data range.

4.D.2 Wasserstein similarity metric

The deconstruction of the data into fPCs and corresponding coefficients enables the use of standard statistical techniques to define measures of similarity between load profiles in the dataset. A load profile similarity metric based on an optimal transport approach is proposed. This metric is based on the approximate Wasserstein distance, or Earth-mover's distance, between the distributions of PC coefficients, which provides a measure of the ease with which one probability distribution can be transformed into the other, and hence is a measure of the similarity of the two distributions. As such, the metric is referred to as the 'Wasserstein similarity metric', and is computed using the [Geomloss](#) (Feydy, 2024) library. Advantages of this method are that it provides a comparison of the underlying patterns in the load profiles that is not excessively distorted by sequence misalignments as is the case for simple metrics like RMSE, and it allows time series of different durations to be compared.

Fig. 4.17 shows the kernel density distributions of the V1 fPC coefficients for Buildings 38 and 49. The Wasserstein distance between these distributions is 0.05. Similarity metric values are computed for comparison of the training datasets between pairs of buildings, and for the comparison of pairs of training, validation, and test datasets for each building individually. Figures 4.18 & 4.19 provide the results of these similarity metric calculations.

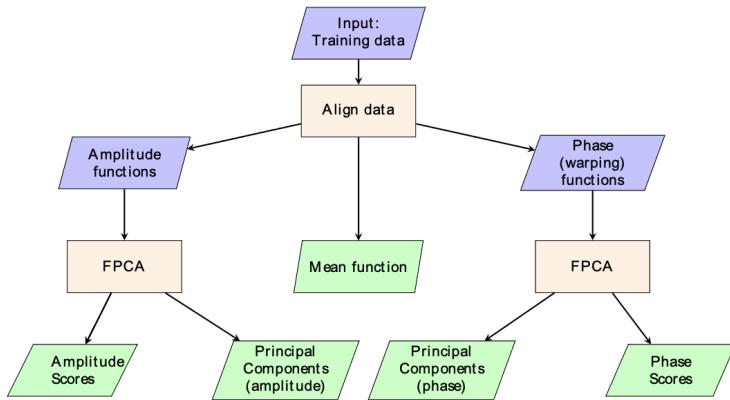


Fig. 4.14 Schematic process for functional principal component analysis (fPCA)

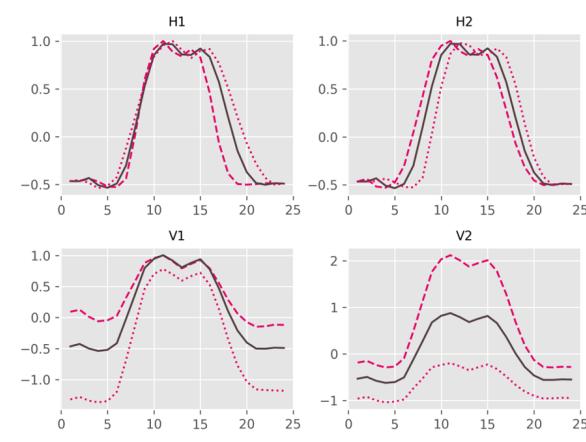


Fig. 4.15 Illustration of first two Phase (H) and Amplitude (V) fPCs. Solid black line shows mean function, $\mu(t)$. Dashed line (-) indicates the impact of a +ve coefficient and dotted line (.), a -ve coefficient.

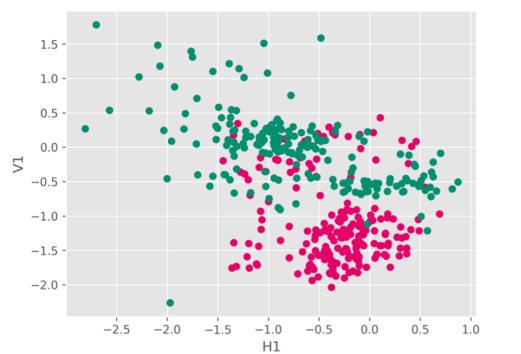


Fig. 4.16 Example fPC coefficient distributions for Buildings 38 (pink) and 49 (green)

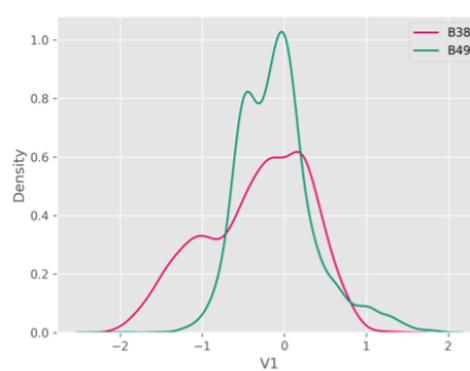


Fig. 4.17 Kernel density plot of V1 scores for Buildings 38 & 49. Wasserstein distance between two distributions is 0.05.

	L0	L3	L9	L11	L12	L15	L16	L25	L26	L32	L38	L44	L45	L48	L49
L0	0.000	1.888	5.037	0.480	1.019	2.368	0.425	0.296	0.255	0.464	1.903	0.344	0.273	0.484	1.817
L3	1.888	0.000	2.744	1.900	1.527	1.651	1.598	1.625	1.894	2.035	1.149	1.819	1.924	1.873	2.150
L9	5.037	2.744	0.000	5.135	4.748	3.138	4.581	4.409	5.290	5.303	2.811	5.149	4.953	5.096	3.413
L11	0.480	1.900	5.135	0.000	0.967	2.455	0.441	0.510	0.432	0.299	1.810	0.395	0.530	0.314	1.831
L12	1.019	1.527	4.748	0.967	0.000	2.378	0.679	0.895	1.046	1.032	1.345	0.784	0.917	0.786	1.417
L15	2.368	1.651	3.138	2.455	2.378	0.000	2.228	2.150	2.692	2.390	2.120	2.454	2.480	2.363	1.892
L16	0.425	1.598	4.581	0.441	0.679	2.228	0.000	0.467	0.472	0.488	1.497	0.441	0.454	0.439	1.400
L25	0.296	1.625	4.409	0.510	0.895	2.150	0.467	0.000	0.350	0.615	1.557	0.312	0.304	0.420	1.575
L26	0.255	1.894	5.290	0.432	1.046	2.692	0.472	0.350	0.000	0.466	1.964	0.321	0.254	0.474	2.144
L32	0.464	2.035	5.303	0.299	1.032	2.390	0.488	0.615	0.466	0.000	1.936	0.490	0.555	0.399	1.871
L38	1.903	1.149	2.811	1.810	1.345	2.120	1.497	1.557	1.964	1.936	0.000	1.683	1.862	1.753	1.735
L44	0.344	1.819	5.149	0.395	0.784	2.454	0.441	0.312	0.321	0.490	1.683	0.000	0.364	0.307	1.755
L45	0.273	1.924	4.953	0.530	0.917	2.480	0.454	0.304	0.254	0.555	1.862	0.364	0.000	0.506	1.717
L48	0.484	1.873	5.096	0.314	0.786	2.363	0.439	0.420	0.474	0.399	1.753	0.307	0.506	0.000	1.583
L49	1.817	2.150	3.413	1.831	1.417	1.892	1.400	1.575	2.144	1.871	1.735	1.755	1.717	1.583	0.000

Fig. 4.18 Similarity metric scores between training datasets for pairs of buildings.

	L0	L3	L9	L11	L12	L15	L16	L25	L26	L32	L38	L44	L45	L48	L49
train_validation	0.213	0.600	0.412	0.215	0.537	0.851	0.633	0.453	0.281	0.184	0.783	0.202	0.245	0.390	0.422
validation_test	0.256	0.721	1.622	0.319	0.328	0.319	0.571	0.345	0.295	0.219	0.518	0.263	0.312	0.151	0.722
train_test	0.249	1.041	1.426	0.362	0.621	0.905	0.508	0.690	0.319	0.244	1.157	0.281	0.266	0.408	0.658

Fig. 4.19 Similarity metric scores between pairs of training, validation, test datasets for each building.

Appendix 4.E Change-point analysis

Building electrical loads are driven by occupancy behaviour dynamics which span timescales: seasonal (winter heating vs. summer cooling loads), weekly (inhabitance trends due to working patterns, i.e. weekday vs. weekend), and daily (nighttime vs. daytime). The form of these behavioural trends also differs between buildings depending on their usage, e.g. offices and residential buildings have opposing daily trends. Gradual changes can occur in these behavioural trends, for instance due to cultural shifts in working patterns. Alternatively, abrupt changes can be caused by severe disturbances, such as building stock refurbishment, temporary vacancy, or occupant behavioural changes due to change in building use. As such, changes can occur in the seasonal and/or trend patterns of building load time series, and the underlying patterns of the dynamics can differ significantly following disturbances.

This raises the question as to whether all historic measurement data is useful for the training of load prediction models, as the underlying patterns in certain sections of the historic data may differ significantly from the current behaviour of the building load, i.e. be non-representative. Including these sections in the training dataset may poison model training, leading to worse prediction performance.

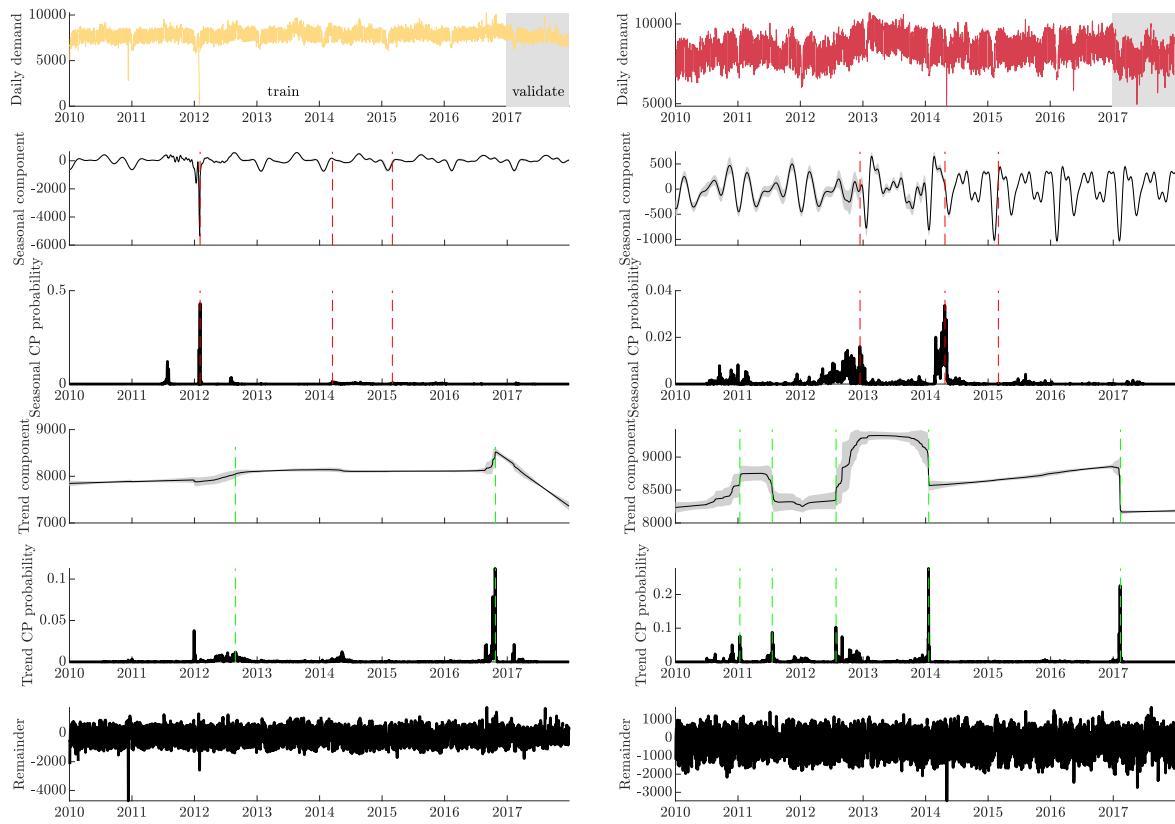


Fig. 4.20 Change-points detected using the BEAST algorithm in load profiles of Buildings 26 & 44 (top row). Signal is decomposed into a seasonal component (second row), assumed harmonic, inferred change-points indicated by red vertical lines, and a trend component (fourth row), with green lines indicating change-points. Credible intervals of the estimated signals are given by grey envelopes around the individual components.

A change-point analysis is conducted on the building electrical load dataset to detect changes in the load patterns in each building. The Bayesian ensemble algorithm “Bayesian Estimator of Abrupt change, Seasonal change, and Trend” (BEAST) (Zhao et al., 2019) is used to decompose the load time series and derive non-linear dynamics across multiple timescales, detecting change-points, seasonality, and trends. The value of this ensemble approach is that it quantifies the relative usefulness of individual decomposition models, leveraging all the models via Bayesian model averaging. Additionally, it provides an estimated probability of inferred change-points being ‘true’ change-points. Results of the change-point analysis for Buildings 26 & 44 are shown in Fig. 4.20. The demand profiles are decomposed into seasonal and trend components, and detected change-points for the two components are indicated in red and green respectively. The trend component of Building 44 is found to vary substantially more than that of Building 26, which corresponds with the visual variability of the demand profiles.

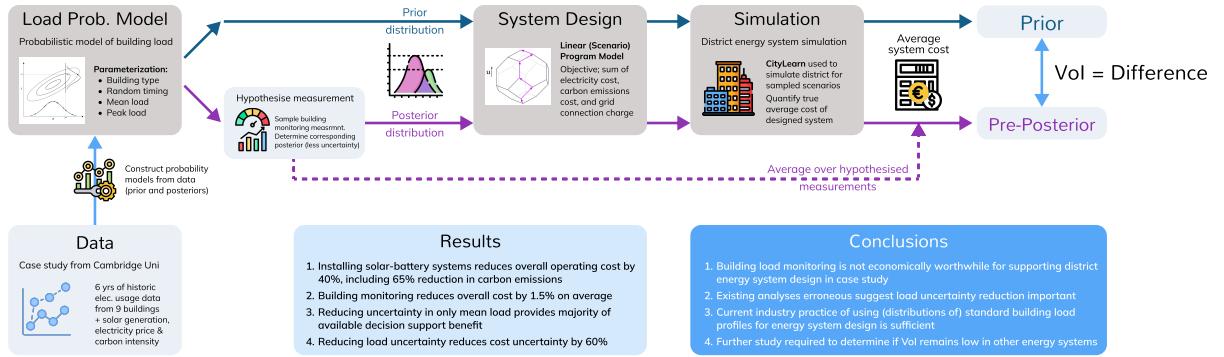
5

Supporting solar-battery system sizing with load monitoring

Max Langtry and Ruchi Choudhary (Dec. 2025a). “Quantifying the Benefit of Load Uncertainty Reduction for the Design of District Energy Systems under Grid Constraints Using the Value of Information”. In: *Applied Energy* 400, p. 126549. ISSN: 0306-2619. DOI: [10.1016/j.apenergy.2025.126549](https://doi.org/10.1016/j.apenergy.2025.126549). URL: <https://www.sciencedirect.com/science/article/pii/S0306261925012796>

This chapter has been published as the journal article above.

All code and data used to perform the experiments in this chapter is available at <https://github.com/mal84emma/Building-Design-Vol>.



Icon credits: Flaticon.com

Fig. 5.1 Graphical abstract of Chapter 5.

When building energy systems are designed there is substantial uncertainty in how much energy the building will actually use, and how that energy usage will look. This load uncertainty needs to be accounted for during design to make sure the building energy system can both meet the building's energy needs and manage its impact on the grid. Reducing uncertainty in building load allows for improved designs with less compromise to be identified, reducing the cost of decarbonising the building's energy usage. However, monitoring the building and gathering the data required to reduce load uncertainty at the time the energy system is designed is costly. This chapter quantifies the economic value of practical monitoring data for improving the design of building energy systems, to determine whether the benefits of reducing load uncertainty outweigh the costs. The On-Policy **Vol** framework from Section 2.5.1 is applied to a case study district energy system design problem. This case study uses a Linear Program model to size solar-battery systems and grid connection capacity under uncertain building loads, which are modelled using historic electricity metering data from the Cambridge University estate. The chapter investigates the impact that the building load uncertainty has on the performance of the district energy system, and the impact that reducing that uncertainty has on the system design.

5.1 Introduction

Many countries such as the UK and EU have implemented policies focusing on the adoption of heat pumps to decarbonise building heating through electrification (House of Commons, 2022; IEA, 2021). However, achieving a widespread roll-out of heat pumps comes with significant challenges for decarbonising building electricity usage and managing the electrical grid. Electrifying heating will substantially increase the total demand for electricity, and more significantly peak electricity demand (Heinen et al., 2018), which would rise by 50% in the UK (Charitopoulos et al., 2023; M. Zhang et al., 2022). Supporting this additional electricity demand while maintaining security of supply requires investment in distribution and transmission systems (Blonsky et al., 2019), as well as reserve generation capacity (Charitopoulos et al., 2023), which add to the cost of heat decarbonisation. Already, capacity limitations on the electrical supply system are causing delays to both building construction and renewable power generation projects (BEIS, 2023; Department for Business and Trade, 2023), and it is widely acknowledged that network capacity limitations are a significant barrier to future decarbonisation (Haben and Young, 2023).

To achieve 2050 decarbonisation targets, energy retrofits must be cheap enough so that they are widely adopted, but also able to be safely integrated into the wider energy system. This means building energy systems must be designed to balance their total system cost, the carbon emissions reductions they achieve, and their impact on the electrical grid. The installation of distributed generation and storage in buildings has become a popular route to reducing the carbon emissions of building energy usage, and so has attracted considerable research attention in recent years (B. Li et al., 2023; Niveditha and Rajan Singaravel, 2022; Novoa et al., 2019; Salvador and Grieu, 2012; Sharma et al., 2020; Tumminia et al., 2020). This is due to the falling costs of solar PV generation (IRENA, 2023), and the need for local storage to both improve buildings' utilization of the solar energy and reduce the grid impact caused by peaks in solar generation (B. Li et al., 2023).

When designing building energy systems, there is substantial uncertainty concerning the conditions in which they will operate. This includes weather conditions which affect local energy generation, building usage type and occupant behaviour which determine the timing of energy usage, and building envelope characteristics which influence the building thermal response. Of particular importance for managing grid impacts is the electricity usage behaviour of a building, including the timing and magnitude of peak loads. Due to the greater weather dependence and seasonal variability of heating loads, electrification of heating will increase uncertainty in building electricity usage (Deroubaix et al., 2021; Egging-Bratseth et al., 2021), in terms of both the aggregate and peak demand. The challenges posed by these uncertainties, and methods for properly accounting for them

during the design of building energy systems have been widely studied (DeCarolis et al., 2017; Fodstad et al., 2022; Mavromatidis et al., 2018a; Tian, Heo, et al., 2018; L. Zhu et al., 2022).

There are two routes to improve the design, and so eventual operational performance, of future decarbonised building energy systems. Firstly, improving optimisation methods for designing energy systems in the presence of uncertainties. Secondly, reducing the uncertainty in the conditions in which the energy system will operate, by for example monitoring the heating and electricity use patterns of the building prior to designing the heating retrofit. Optimisation methodologies have received concerted research effort for many years (Connolly et al., 2010; Fodstad et al., 2022), whereas there has not yet been any study of the potential benefits of uncertainty reduction for improving energy system design.

Reducing uncertainty in the operational conditions of a building improves energy system design by allowing a design that is better tailored to actual needs of the building to be identified, i.e. one with less hedging/compromise. So uncertainty reduction can lower the cost of decarbonising the building's energy usage. However, uncertainty reduction is itself costly. For the case of existing buildings, these costs arise from the hardware and software needed to monitor energy consumption patterns, and the cost of management, curation, analysis, and interpretation of the data. In the UK, only 51% of electricity and gas meters are currently 'smart' (able to collect hourly energy usage data), despite rollout beginning in 2012, and expected to cost £19.4bn in total (DESNZ, 2023f). So it is important to quantify the benefit that uncertainty reduction via monitoring buildings provides to energy system design. Specifically, whether the reductions in total system cost are greater than the cost of gathering and exploiting the data needed to reduce uncertainty.

5.1.1 Quantifying the impacts of uncertainty on building energy system design

As discussed in the [Introduction](#), the impact that uncertainties in buildings have on the performance of energy systems, and methods for accounting for and managing these uncertainties during system design, have been studied extensively (see Section 1.2.2). However, there has been very little study of the impact these uncertainties have on the process of designing building energy systems.

A few previous studies have investigated the influence uncertainties have on the optimal system design. (Yongjun Sun, 2015) analyses the design of a Net Zero Energy Building with distributed generation and storage, and plots how the capacities of each system component (chillers, wind turbines, solar panels, and battery) and the total cost of the optimized system change with uncertain values of properties of the building (including wall

thickness, internal heat gain, temperature set point). Additionally, influence coefficients of the uncertain parameters are computed for the capacity of each system component. A later study, (Sanajaoba Singh and Fernandez, 2018), performed a similar analysis for another system with distributed generation and storage, investigating how the optimal design and total cost changed with varying scenarios of renewable generation. In both cases no statistical distributions of the uncertain parameters were considered, and so the insights gained are limited. For example, in the case of (Sanajaoba Singh and Fernandez, 2018), it is unclear whether a design with a substantially greater number of PV panels is likely to be optimal, and so worth investing in.

A statistical [Sensitivity Analysis](#) for the optimal design of an energy system with distributed generation and storage was performed in (Mavromatidis et al., 2018b). In this study, a Mixed Integer Linear Program was used to optimize the sizing of boilers, heat pumps, thermal storage, and PV generation for scenarios defined by samples from distributions of uncertain parameter values, including energy and capital costs, conversion efficiencies, and energy demand. This enabled a distribution of optimal sizes to be created for each component to illustrate how the uncertainties affect the optimal design choice. (Pickering and Choudhary, 2019) performed a similar procedure for the sizing of components in a district energy system, however the optimized designs were not analyzed. Whilst these studies identify the distribution of optimal designs resulting from the underlying uncertainties in the system, they do not then provide a way to quantify the importance of those uncertainties for the design process.

[Value of Information](#) analysis provides a clear and rigorous methodology for quantifying the benefit of uncertainty reduction for improving decision making (see Chapter 2). However so far only one previous study has applied it to the design of building energy systems. This study, (J. Niu et al., 2023) which is discussed in more detail in Section 1.2.3, quantifies the reduction in the total cost of a district energy system that would be achieved if the building energy demands and solar generation could be perfectly known when sizing the system components (including chillers, solar panels, and energy storage), i.e. if all uncertainty were removed. But this can't be achieved by any practical measurement, and so the results of this study can't inform us of the usefulness of practical building monitoring to support energy system design.

5.1.2 Research objectives & contributions

There has been very limited study of the economic benefit of uncertainty reduction from measurement/data collection to support decision making in building energy systems. No existing study has quantified the benefit of using hourly energy metering data to support the design of district energy systems by reducing uncertainty in building load profiles. Understanding of the value of load monitoring for improving energy system design is needed to determine whether significant cost savings are currently being missed, or whether current industry practice of using standard load profiles for design is sufficient.

This chapter uses the On-Policy [Vol](#) framework to quantify the cost savings achieved when using hourly building monitoring data to improve energy system design. It investigates the sizing of solar-battery systems and grid connection capacity for an illustrative district energy system, where probabilistic models of building load profiles are constructed using historic energy usage data from the Cambridge University estate. The main research objectives are to:

- Determine if load monitoring is economically worthwhile for improving energy system design by reducing load uncertainty;
- Quantify the relative importance of load profile features (mean load, peak load, and shape) for improving system design; and
- Gather evidence to assess whether current industry practice of using standard load profiles for building energy system design is sufficient

The key research contribution is that this work is the first to quantify the economic value of using practical building measurement to support energy system design through load uncertainty reduction. The analysis in this chapter demonstrates that existing methods for assessing the impact of uncertainties on energy system design are insufficient, and can lead to erroneous conclusions about the importance of uncertainty reduction. The results provide the first numerical evidence to indicate that existing industry practice of using (distributions of) standard building load profiles for energy system design is sufficient for minimizing the cost of building energy decarbonisation.

The remainder of this chapter is structured as follows. Section [5.2](#) describes the case study district energy system. Sections [5.2.1](#) and [5.2.2](#) present the data studied, and how it is used to construct probabilistic models of building loads. Next, Sections [5.2.3](#) and [5.2.4](#) define how the system is designed using a Linear (Scenario) Programming model, and evaluated via simulation. The [Vol](#) calculations are performed in Section [5.3](#), and the practical importance of the results obtained is discussed. Finally, conclusions are drawn in Section [5.4](#).

5.2 Case study district energy system

The benefits of uncertainty reduction to support energy system design are investigated for a problem involving the sizing of solar-battery systems to decarbonize electricity usage in a grid-constrained district energy system with electrified heating. At the time of design, the patterns and scale of electrical load in the buildings of the district are unknown/uncertain. This load uncertainty must be managed during design to allow the district to operate within the grids constraints. Specifically, the peaks in electrical load must be manageable, though their magnitude and timing (particularly the co-incidence of peaks) are not known during design.

Prior to committing to a system design (PV, battery, and grid connection sizes), the designer may choose to gather hourly electricity usage for the buildings from smart meters, reducing uncertainty in both the patterns and scale of their electrical load. This enables a more tailored design with a lower total cost to be identified. However, gathering hourly metering data for each building in the district and applying it to size the solar-battery systems is costly. This case study asks the question,

Does building monitoring to reduce uncertainty in electrical load provide a net economic benefit when designing solar-battery systems?

i.e.

Does the reduction in cost from improved design outweigh the costs of collecting hourly monitoring data needed to achieve uncertainty reduction?

5.2.1 Case study data

Historic building energy usage data from the Cambridge University Estates dataset ([Langtry and Choudhary, 2024](#)), described in Appendix A, is used to provide a case study for the design of a district of new academic buildings⁴⁹.

From this dataset, 9 buildings with available electricity and gas usage data for the period 2012 to 2017 are selected (the longest period in the dataset with a significant number of buildings with high quality data available). For each building, data for each year is separated out. To model energy usage in buildings with electrified heating, the electricity

⁴⁹ At the time of writing, such a development project is under consideration, and one significant hurdle for its implementation is the capacity restrictions on the local power distribution network, particularly as heat pumps are a leading option for minimising heating related carbon emissions.

and gas usage measurements are combined to produce overall electrical load profiles such that equal amounts of electricity are used by direct electricity and heating (meaning that assuming a SCOP of 3, typical for ASHPs in the UK (Terry and Galvin, 2023), the direct electricity to final heating energy usage matches the current 25:75 ratio (EEA, 2023)). Fig. 5.2 plots examples of these building electrical load profiles (normalized by their mean loads) for periods in the summer and winter. It demonstrates the variation in electricity usage behaviour both between buildings, and between years for a given building.

Data from 2023 is used for grid electricity price and grid carbon intensity, as these provide the closest available representation of the conditions the district will face during operation. To represent the uncertainty in solar generation, years of normalised solar generation data are sampled randomly for each scenario from a dataset of the years 2010 to 2019. For the case study, these data are treated as exogenous, and are used to define the probabilistic operational conditions of the energy system.

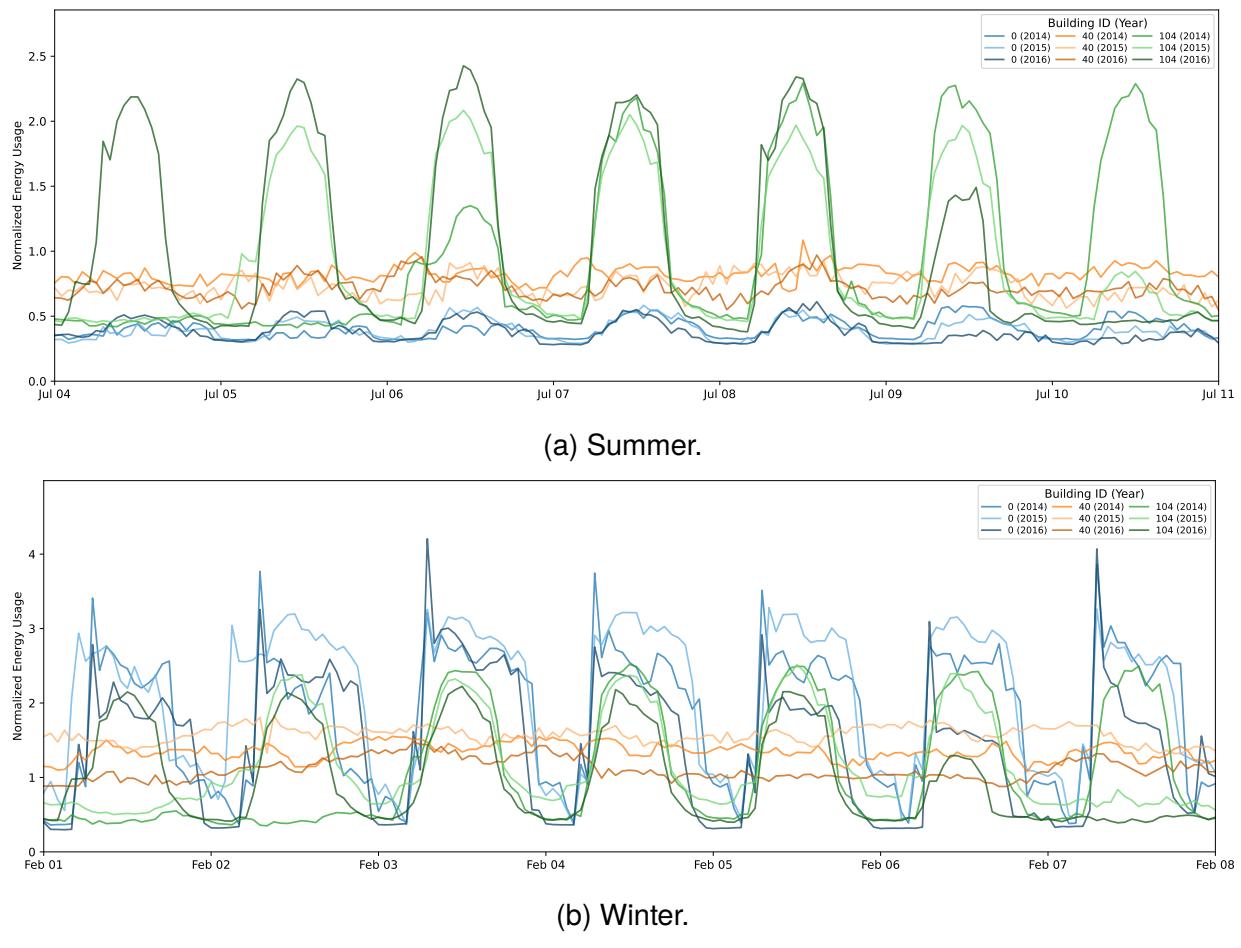


Fig. 5.2 Example electricity load profiles, representing buildings with electrified heating.
The complete set of electrified building load profiles considered in the case study can be view interactively [here](#).

5.2.2 Probabilistic models of uncertain building electrical load & measurement via monitoring

A probabilistic model of the uncertain electricity load of a building is constructed from the data. The functional data variables representing building load are parameterized to provide a simple statistical representation that allows the importance of different characteristics of electrical load to be studied. Four parameters are used to specify the load profiles, two representing the shape of the profile (`type` and `year`), and two representing the scale of electricity usage (`mean` and `peak`).

The probabilistic models of profile shape are taken directly from data. The parameter `type` is used to represent the broad patterns of energy usage exhibited by a building, which derive largely from its usage type, e.g. laboratory, teaching, office, or a mix thereof. It is taken as a categorical variable denoting which building in the case study dataset the load profile data is taken from. All buildings are assumed equally likely,

$$\text{type} \sim \mathcal{U}(\{\text{available buildings}\}) \quad (5.1)$$

The parameter `year` is used to represent the inherent randomness in building energy usage due to factors that cannot be known in advance, such as occupant behaviour and weather. It is taken as a categorical variable denoting the year from which the load profile data is taken, which are all assumed equally likely,

$$\text{year} \sim \mathcal{U}(\{2012, \dots, 2017\}) \quad (5.2)$$

Load monitoring is assumed to provide knowledge of the broad energy usage patterns of a building, i.e. perfect information for the parameter `type` and no information for `year`.

Probabilistic models for the scale of electricity usage are determined from the statistics of the available load data. The prior model of mean building load is taken to be,

$$\text{mean} \sim \mathcal{N}(100, 25) \quad (5.3)$$

representing a $\pm 50\%$ uncertainty in a design load of 100kW.

The prior model of peak building load is taken to be,

$$\text{peak} \sim \mathcal{U}(200, 400) \quad (5.4)$$

which is fit using the annual peak loads from the dataset.

In a practical building, load monitoring provides an uncertain/imperfect estimate of the mean and peak building load, as there is year-on-year variation in these values that

cannot be predicted in advance. The likelihood models for the precision of mean and peak load measurements (estimates) from building monitoring are taken to be Gaussian, with measurement errors determined using the year-on-year variations observed in the load dataset, i.e. the likelihood of obtaining a measurement value is,

$$z|\theta \sim \mathcal{N}(\theta, \varepsilon\theta) \quad (5.5)$$

with $\varepsilon = 0.1$ for mean, and $\varepsilon = 0.075$ for peak.

The prior probabilistic models of the parameters describing the building electricity load profiles, and the likelihood models for measurements made via building monitoring are summarized in Table 5.1. Where necessary, the corresponding posterior distributions are modelled using Stan (Carpenter et al., 2017).

The probabilistic models of electrical load are taken to be the same for all buildings in the district, with the parameters type, mean, and peak independent for each building. The year parameter is the common to all buildings (sampled once for the district in each scenario). This is done to ensure that correlations in the timing of energy usage between buildings are properly captured, as building energy usage is driven by common externalities across the district, for example outdoor temperature which affects heating load. These correlations in building loads are important for correctly representing the overall energy usage behaviour of the district. For example, the relative timing of peaks in building load determines the district peak load, and so the required grid connection capacity.

Samples are drawn from the building load distributions (either prior or posterior) by sampling the four parameters (type, year, mean, peak) using their probabilistic models, retrieving the constructed electrified heating load data for the sampled building (type) and year (see Fig. 5.2), and scaling the profile to the sampled mean and peak load values.

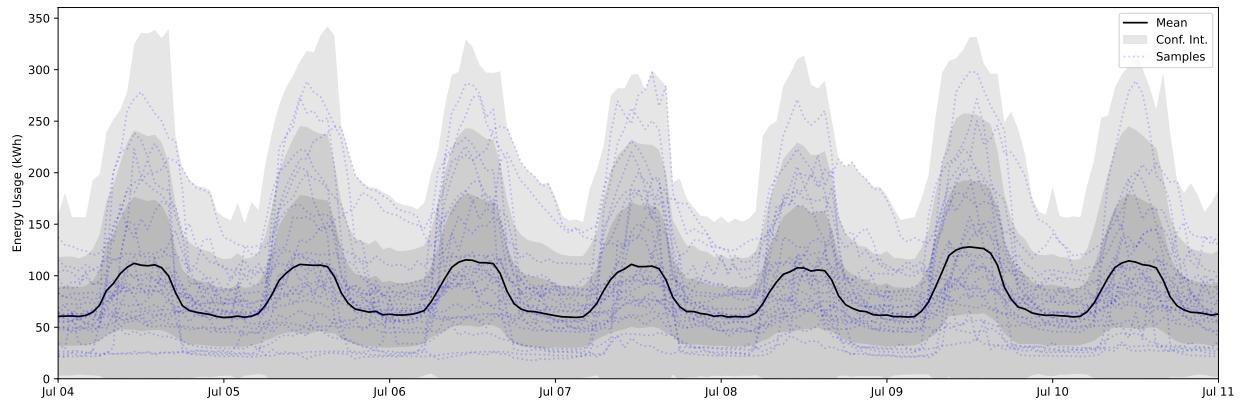
Parameter	Prior model	Prior params	Measurement model	Msr. params
type year*	Discrete uniform	{building ids} {data years}	Exact (perfect info) None (no info)	– –
mean	Gaussian	$\mu = 100 \text{ kW}$ $\sigma = 25 \text{ kW}$	Gaussian imprecision (imperfect info)	$\varepsilon = 0.1$
peak	Continuous uniform	min = 200 kW max = 400 kW	$z \theta \sim \mathcal{N}(\theta, \varepsilon\theta)$	$\varepsilon = 0.075$

Table 5.1 Probabilistic models of parameters describing uncertain building load, and their measurement via monitoring.

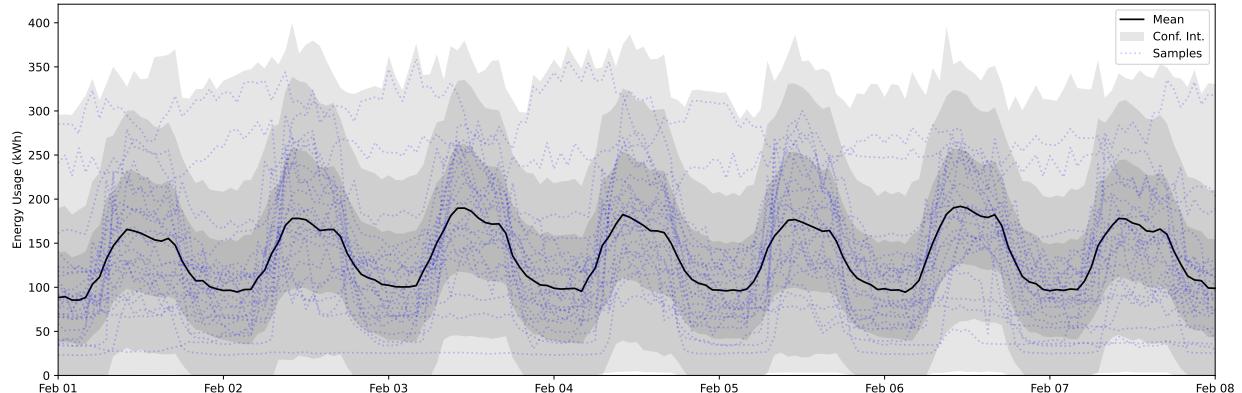
*Parameter common to all buildings in the district

The prior probabilistic model of building electrical load is visualised in Fig. 5.3. The mean load profile is indicated in black, with confidence intervals of ± 1 s.d., ± 2 s.d., and the range (min to max), indicated in successively lighter shades of gray. A set of load profiles sampled from the distribution are traced in blue.

Fig. 5.4 illustrates the posterior distribution of building load for an example load profile measurement obtained by building monitoring (indicated in red). Load profiles sampled from the posterior distribution are traced in green. The confidence intervals of the prior distribution, from Fig. 5.3, are underlaid for comparison. The tighter clustering of the profiles sampled from the posterior compared to that for the prior samples in Fig. 5.3 demonstrates the reduction in building load uncertainty achieved by monitoring.



(a) Summer.



(b) Winter.

Fig. 5.3 Visualisation of prior probabilistic model of uncertain building load, with profiles sampled from distribution.

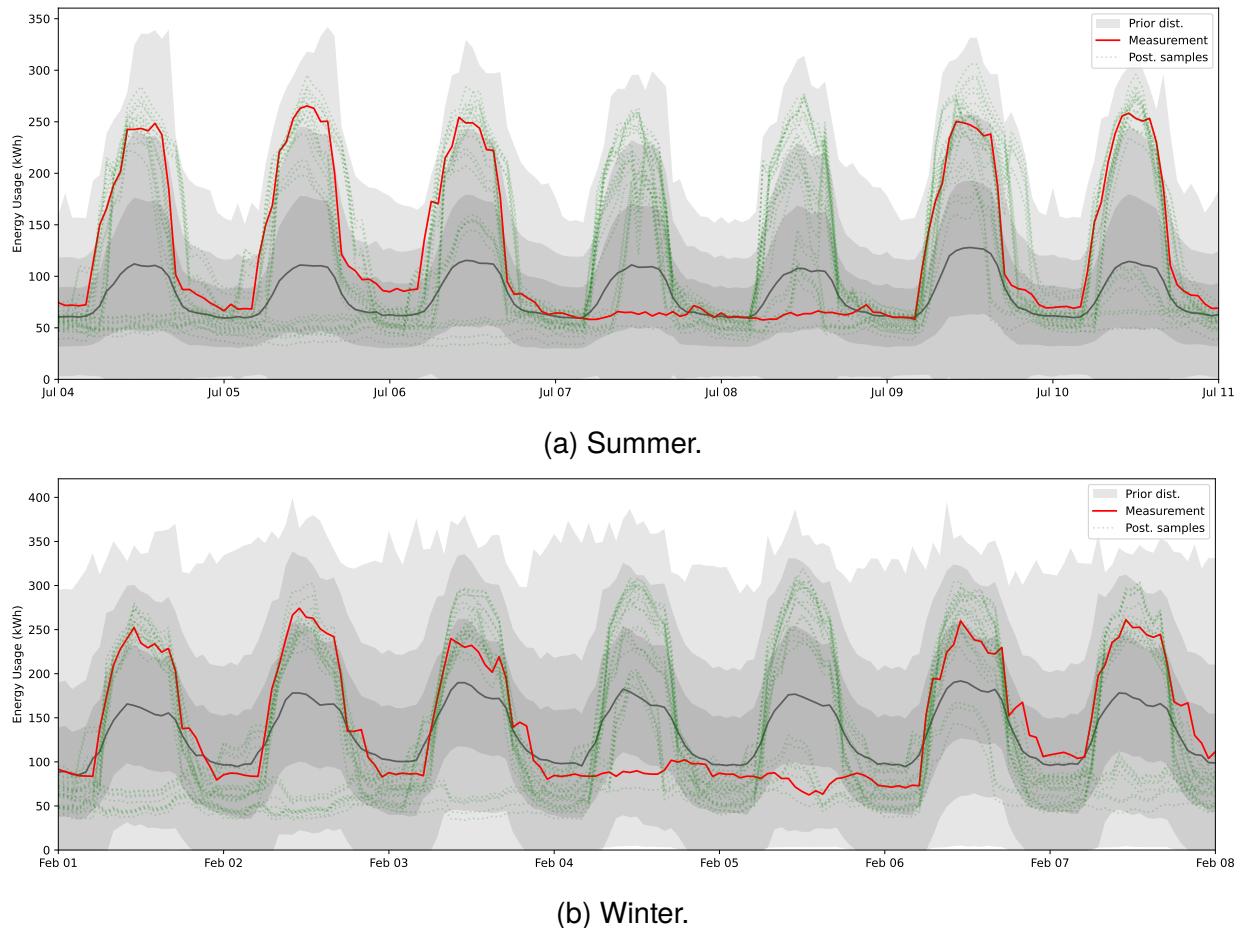


Fig. 5.4 Visualisation of posterior probabilistic model of building load for example measurement.

5.2.3 Stochastic Programming for system sizing

For the case study, Stochastic Programming is used as a policy to optimize the sizing of solar PV, battery storage units, and grid connection capacity, to determine district energy system designs that (attempt to) minimize the expected total system cost, accounting for uncertainty in electrical load and solar generation. Stochastic Programming (specifically linear scenario programming) is chosen as the modelling and optimization methodology due to its computational efficiency for problems with large numbers of decision variables, and its resulting prevalent use in the current literature for designing district- and national-scale energy systems (DeCarolis et al., 2017; *Open Energy Modelling Initiative Wiki* 2024; Pickering and Choudhary, 2019; Yun Yang et al., 2015; Yue et al., 2018).

A linearized model of the electrical behaviour of the district energy system is used. A schematic of the energy flows within the district energy system model is provided in Fig. 5.5. This allows the design optimization task to be formulated as a Linear Program, which is presented in Eq. 5.6, with Table 5.2 providing descriptions of the parameters. The model formulation uses the same form of constraints to represent energy system behaviour as state-of-the-art models such as (Pfenninger and Pickering, 2018) and (Brown, Hörsch, et al., 2018).

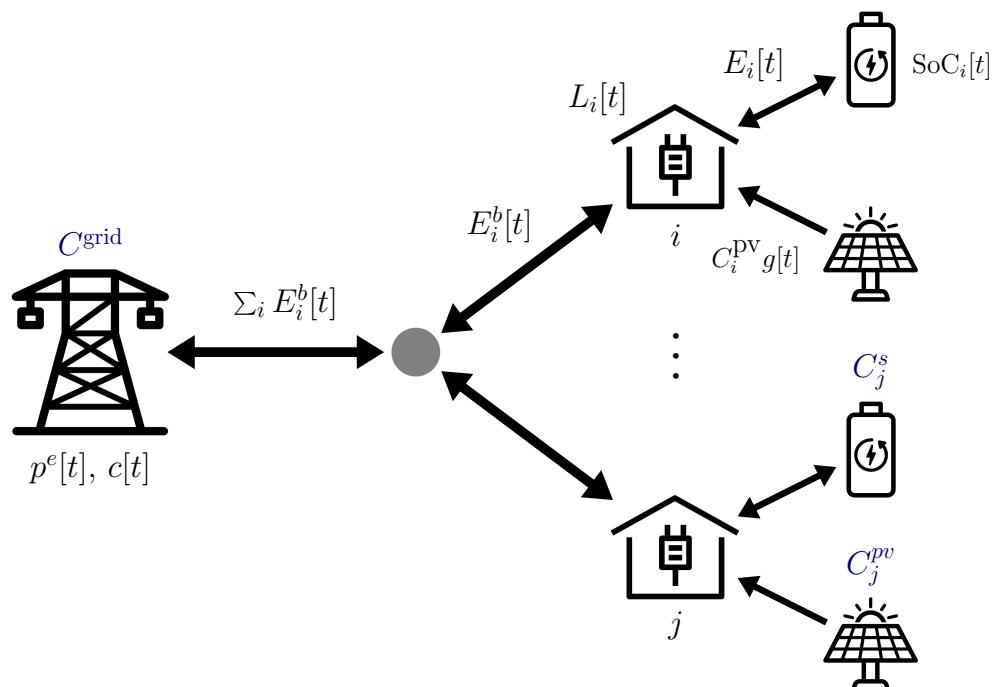


Fig. 5.5 Energy flow schematic for case study district energy system.

Operational variables are indicated in black, capacity variables are indicated in blue. (Icon credits: [Symbolon](#))

$$\begin{aligned}
\min \quad & \gamma \sum_m \rho_m \left(\underbrace{\sum_t \left(p^e[t] \sum_i \left(\max \left[0, E_{i,m}^b[t] \right] \right) + p^c c[t] \max \left[0, \sum_i E_{i,m}^b[t] \right] \right)}_{\text{electricity cost}} \right. \\
& \left. + \underbrace{p^{\text{grid excess}} \max \left[0, \max_t \left[\left| \sum_i E_{i,m}^b[t] \right| \right] / \Delta t - C^{\text{grid}} / \text{FoS} \right]}_{\text{grid excess cost}} \right) \\
& + \underbrace{\sum_i \left(p^s C_i^s + p^{\text{pv}} C_i^{\text{pv}} \right)}_{\text{capital cost}} + \underbrace{\gamma p^{\text{grid}} C^{\text{grid}}}_{\text{grid connection cost}}
\end{aligned} \tag{5.6}$$

over $C_i^s, C_i^{\text{pv}}, C^{\text{grid}}, E_{i,m}[t], \text{SoC}_{i,m}[t+1] \quad \forall i, m, t$

subject to $\text{SoC}_{i,m}[t+1] = \text{SoC}_{i,m}[t] + \sqrt{\eta_i} [E_{i,m}[t]]^+ - 1/\sqrt{\eta_i} [E_{i,m}[t]]^- \tag{5.6a}$

$- P_i^{\max} \Delta t \leq E_{i,m}[t] \leq P_i^{\max} \Delta t \tag{5.6b}$

$0 \leq \text{SoC}_{i,m}[t+1] \leq C_i^s \tag{5.6c}$

$C_i^s, C_i^{\text{pv}}, C^{\text{grid}} \geq 0 \tag{5.6d}$

$P_i^{\max} = \delta C_i^s \tag{5.6e}$

$\text{SoC}_{i,m}[0] = \text{SoC}^0 C_i^s \tag{5.6f}$

$E_{i,m}^b[t] = L_{i,m}[t] - C_i^{\text{pv}} g_m[t] + E_{i,m}[t] \tag{5.6g}$

for all $i \in [0, B-1], t \in [0, T-1], m \in [0, M-1]$

Parameter	Units	Description
Decision variables		
C_i^s	kWh	Energy capacity of battery unit in building i
C_i^{pv}	kWp	Peak power capacity of solar PV unit in building i
O^{grid}	kW	Contracted power capacity of grid connection
$E_{i,m}[t]$	kWh	Energy <i>intake</i> to battery unit in building i at time t in scenario m
$\text{SoC}_{i,m}[t]$	kWh	State-of-charge of battery unit in building i at time t in scenario m
Derived variables		
$E_{i,m}^b[t]$	kWh	Net energy flow <i>into</i> building i at time t in scenario m
P_i^{\max}	kW	Power capacity of battery unit in building i
Parameters (data)		
γ	yrs	System lifetime
Δt	hrs	Time step of simulation data
ρ_m	–	Probability of scenario m
δ	kW/kWh	Discharge ratio of battery units (power capacity/energy capacity)
SoC^0	–	Initial state-of-charge of battery units (fraction of capacity)
η_i	–	Round-trip efficiency of battery unit in building i
$L_{i,m}[t]$	kWh	Electrical demand of building i at time t in scenario m
$g_m[t]$	kW/kWp	Normalised generation power from solar PV at time t in scenario m
$p^e[t]$	£/kWh	Grid electricity price at time t
p^c	£/kgCO ₂	Nominal carbon price
$c[t]$	kgCO ₂ /kWh	Carbon intensity of grid electricity at time t
p^s	£/kWh	Unit price of battery storage
p^{pv}	£/kWp	Unit price of solar PV
p^{grid}	£/kW	Price of contracted grid connection capacity (over operational period)
$p_{\text{excess}}^{\text{grid}}$	£/kW	Excess charge for exceeding contracted grid capacity (over operational period)
FoS	–	Factor of safety for grid connection capacity

Table 5.2 Description of Stochastic Program variables & parameters.

The design optimization seeks to minimize as its objective the average total lifetime system cost over a set of M scenarios for the district of B buildings, operated for period of length T . This total system cost is comprised of: the capital cost of solar PV and battery units in the buildings, the cost of electricity purchased from the grid (metered per building), the cost of carbon emissions associated with used grid electricity (for the overall district), and the cost of grid usage. This optimization is subject to energy conservation⁵⁰ (Eq. 5.6a) and capacity (Eq. 5.6b & Eq. 5.6c) constraints for operation within each scenario m .

The cost of grid usage is modelled following the form of Distributed Use of Service (DUoS) connection capacity charges, which are the costs charged to large electricity consumers connected to the distribution network in the UK (Acha et al., 2022; Economics, 2022). These consumers must request a ‘contracted’ connection capacity from the distribution operator, for which they are charged a cost per unit of connection capacity (£/kVA). If during operation a consumer draws power from the grid exceeding this contracted capacity, they are charged an additional cost for each kVA that their peak power draw exceeds the contracted capacity. Exaggerated grid charges (compared to current values) are used to represent the pressure on network connection capacity caused by future heat electrification.

As Linear Programs simultaneously optimize over all operational variables, they assume that the district energy system is operated with perfect foresight. This leads to underestimation of the system operational costs, as the battery units within the modelled energy system are scheduled more effectively than could be achieved in practice, reacting to and avoiding future costs which could not be foreseen. This is particularly important for the grid excess costs which depend on peak grid load values that are greatly influenced by the perfect foresight assumption. To account for this underestimation of grid usage in the design model, a factor of safety (FoS) is introduced into the grid excess cost. This FoS was calibrated using preliminary experiments, and was found to greatly reduce the overall cost of designs, as evaluated using the operational simulations introduced in the following section.

The Stochastic Program requires samples (scenarios) to be drawn from the probability space of electrical load profiles for each building in the district. As a result, the space of possible load scenarios is extremely large. However, due to the cubic computational complexity of Linear Programming, only a small number of scenarios can be tractably included in the Scenario Optimization. Therefore, after load profile samples are drawn, scenario reduction (Gioia, 2023; Heitsch and Römisch, 2003) is performed to select a reduced set of 10 scenarios⁵¹ that are statistically representative of the initial sample. A large initial sample must be considered to adequately cover the probability space of district loads. Therefore state-of-the-art scenario reduction methods which optimize each scenario

⁵⁰ $[\cdot]^+$ and $[\cdot]^-$ represent the positive and negative parts of the argument respectively.

⁵¹ The maximum number which could be used by the Stochastic Program given the available computational budget.

individually, and use the resulting distribution of costs as the reduction metric (Pickering, 2019) are intractable. Instead, computationally cheap scenario metrics must be used. The mean, maximum, and standard deviation of the aggregate electrical load for the district are chosen as the metrics for scenario reduction, as simple features representing the shape and scale of the district electricity load.

The parameter values used for all experiments are detailed in 5.A.

5.2.4 Evaluation of operational cost of system designs via simulation

The objective of the Stochastic Program (SP) provides only an approximation of the expected (mean) total cost of the district energy system. The perfect foresight assumption of the Linear Program model causes underestimation of operating costs, a modelling error which is only partly corrected by the factor of safety. The small number of scenarios which can be considered in the Scenario Optimization due to computational limitations causes statistical error in the estimation of the mean cost. The combined model and statistical errors could lead to significant inaccuracy in the cost estimates⁵², and so the Vol. To provide more accurate estimations of the true cost of system designs (the decision maker's target utility), the operation of the district energy system is simulated.

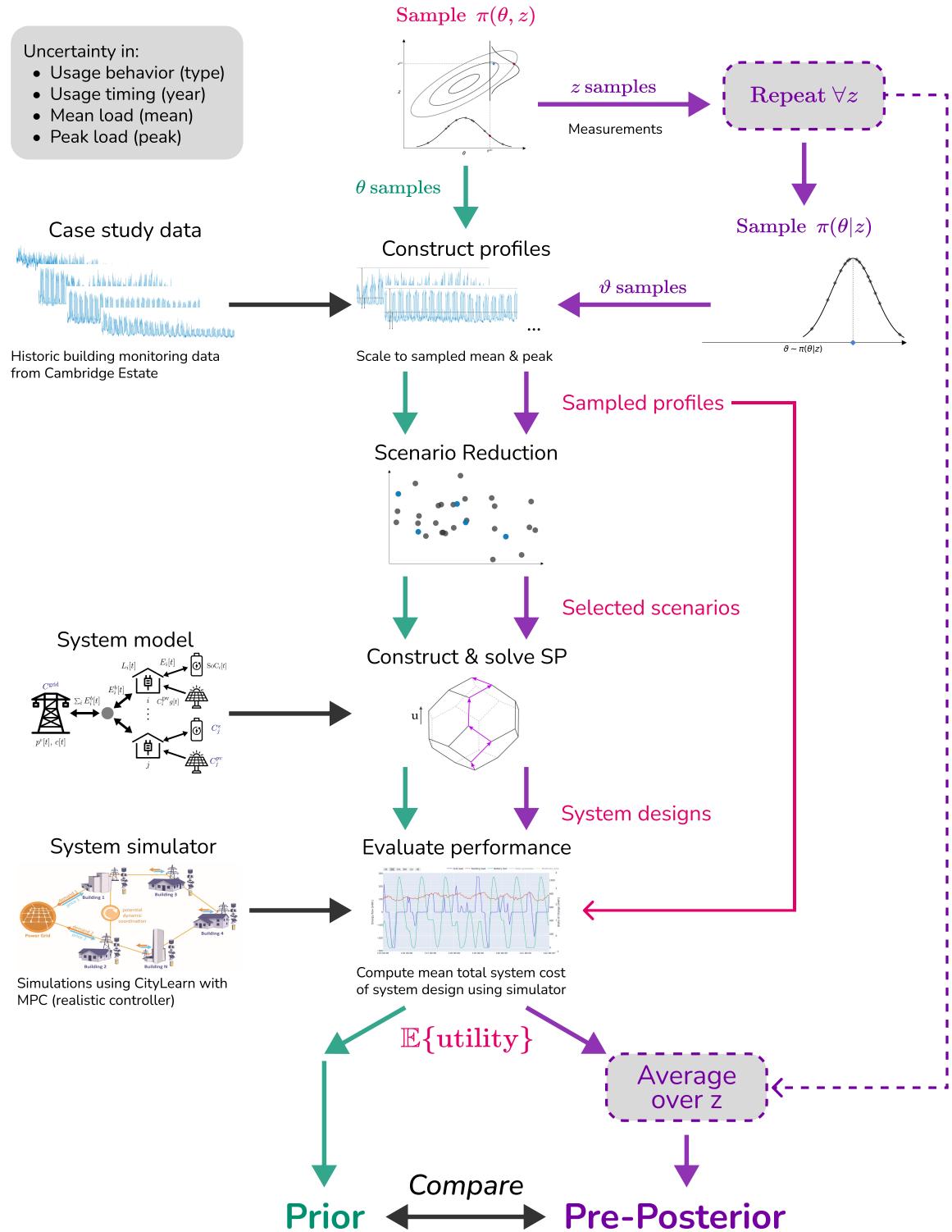
The CityLearn (Vazquez-Canteli, Dey, et al., 2020; Vázquez-Canteli et al., 2019) building energy control framework is used to simulate the behaviour of the district energy system. These simulations provide a more accurate estimate of the operational cost of a system design for a given simulated load scenario. A Model Predictive Control (MPC) scheme is used to schedule battery operation in the simulations. This MPC scheme uses a Linear Programming (LP) formulation with the same form as used for design (Eq. 5.6), however it considers only a single scenario (that being simulated) with perfect forecasts of operational conditions over a 48hr planning horizon. Additionally, a small FoS is used to prevent numerical errors from the LP solver leading to erroneous exceedance of the grid capacity, calibrated from initial experiments. The simulator is configured with linearized building dynamics, such that the MPC scheme perfectly models the system simulation, and so provides perfect system control with 48hrs of foresight, mirroring the SP used for design, but relaxing the perfect foresight assumption. As a result, the simulations still underestimate operational costs (Langtry, Wichitwechkarn, et al., 2024), but this model error is lower than for the SP. As the simulations are computationally cheap, a large number of load scenarios can be simulated, and so the expected cost can be estimated using a large number of Monte Carlo samples (scenarios sampled from the building load distribution), reducing statistical error.

⁵²The accuracy of Stochastic Program objective's estimate of total system cost is investigated in 5.C.2.

5.2.5 Decision problem formulation

The case study can be contextualized within the On-Policy [Vol](#) framework proposed in Section [2.5.1](#) as follows.

The uncertain parameter in the district energy system, θ , is defined as the electrical load profile (functional variable) in each building, and its probabilistic model, $\pi(\theta)$, is described in parametric form in Section [5.2.2](#) (see Table [5.1](#)). The likelihood function, $f(z|\theta)$, describing the uncertainty reduction (imperfect parameter estimates) obtained from building load monitoring, is also described. A Stochastic Programming model, outlined in Section [5.2.3](#), is used as a policy, \mathcal{P}_u , determining a system design (action) to implement given samples from a distribution of building load profiles (to which scenario reduction is applied due to computational limitations). This policy minimizes an approximation of the expected total system cost (utility). The true system cost (utility, u) for a given design is evaluated using simulations of the district energy system with a more realistic control scheme, described in Section [5.2.4](#). Fig. [5.6](#) provides a diagrammatic overview of the methodology.

Fig. 5.6 Methodology for computing Vol for case study district energy system design task.

5.3 Results & Discussion

5.3.1 System design without load monitoring

A district containing 5 university buildings, without constraints on asset sizing, is considered as a base case. Initially the case where no building load monitoring is used is considered. Load profile scenarios were sampled from the prior distribution, which represents the behaviour of a generic university building, and the Stochastic Program model described in Section 5.2.3 was used to size the solar-battery systems and grid connection capacity using these prior scenarios.

The optimized (prior) design installs 4,910 kWh of batteries and 2,790 kWp of solar panels across the 5 buildings, roughly evenly distributed, and contracts 1,230 kW of grid connection capacity (2.5 times the mean electrical load of the district of 500 kW). From operational simulations, the average lifetime cost for the system (the prior expected utility) was determined to be £22.432m over a 20 year operational lifetime, or £1.122m/year. Comparing to the case where no solar-battery systems are installed in the buildings, which results in a mean operational cost of £1.839m/year, this investment in local generation and storage reduces total operating costs for the energy system by 39%. Additionally, the embodied carbon emissions of providing energy are reduced by 64%, and the required grid connection capacity is reduced by 27%. Further, the practical building controller did not exceed the contracted grid capacity in any of the 256 sampled scenarios simulated. These results validate the ability of solar-battery systems to simultaneously reduce the overall cost of energy provision, embodied carbon emissions, and the grid impact of the district. Further details of the optimized system design and its performance across the simulated scenarios are provided in 5.B.

Fig. 5.7 plots the distributions of simulated system costs and corresponding Levelized Costs of Energy (LCOEs) when the prior design is operated for each sampled load scenario. It shows that the uncertainty in building load causes significant uncertainty in the total operating cost of the system. As the LCOE normalizes the operating cost by the total energy usage of the buildings, it removes the contribution of mean load to cost variability. The standard deviation of the LCOE distribution is 3.8%, compared to 10.2% for the total cost distribution, comparing Fig. 5.7b to Fig. 5.7a. This demonstrates that mean load is the greatest driver of cost variability, but that the remaining uncertainties (peak load and load profile shape) also make a significant contribution. There is found to be no correlation between the peak electrical load of the district and the total operating cost across all scenarios, indicating that peak load has little to no influence on the cost of operation. This can be justified as the sizing of the battery units to perform bulk energy

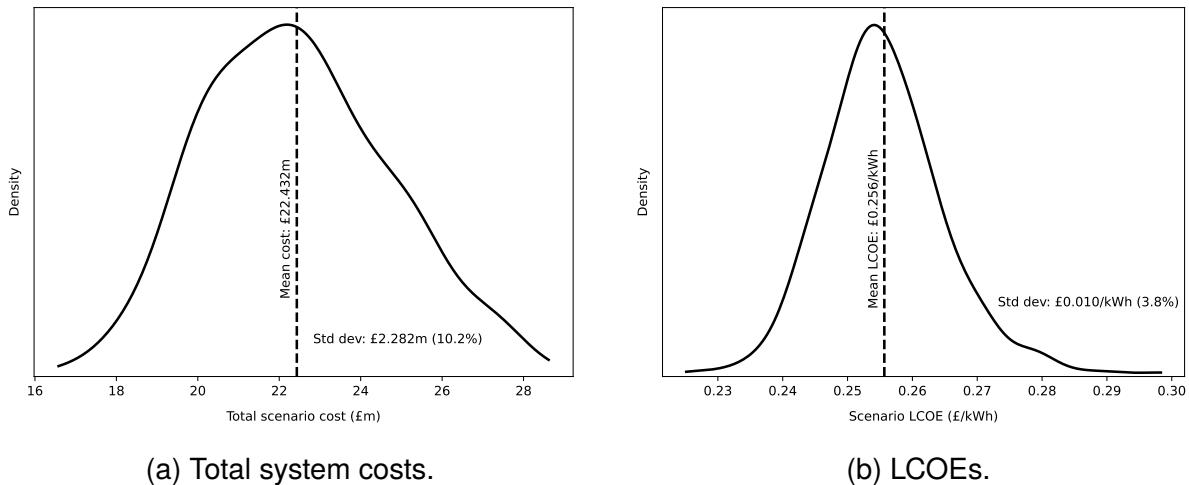


Fig. 5.7 Distribution of simulated total system costs and LCOEs for prior system design over sampled building load scenarios.

arbitrage results in them having a power capacity of 1,960 kW, meaning they can handle the highest possible peak load of the district of 2000 kW (of which only 770 kW needs to be shaved to avoid exceeding the grid connection capacity). As a result, in almost all simulated scenarios no grid excess charges are incurred. Though the peak may occur when electricity is expensive, as they are infrequent there is only a small impact on the overall electricity cost. The operational behaviour of the district energy system and the factors influencing system design are discussed further in 5.B.

This analysis demonstrates that uncertainty in mean electrical load of the district is the largest driver of operational cost uncertainty, responsible for over half of induced variability, but that the other uncertainties (in peak load, load profile shape, and solar generation) also have a significant influence on cost. Therefore there might exist scope to reduce the total system cost by reducing uncertainty in these factors, and tailoring the system design to the energy usage behaviour of the true target building⁵³. Most existing analyses, such as those in (Mavromatidis et al., 2018a; Prataviera et al., 2022), or other purely statistical approaches (Pang et al., 2020), stop at this stage and put forward the conclusion that reducing uncertainty in the model's inputs (here, building load) would significantly reduce uncertainty in the model output (the operating cost of the system). However, this outcome does not provide any indication about whether this uncertainty reduction would affect the decision made, and so cannot quantify its importance for supporting decision making.

⁵³ This tailoring of the system (improved design) could reduce not only the highest cost scenarios for the prior case (the right tail of Fig. 5.7a), but also costs in the scenarios which are already cheap, as the system design does not need to compromise to keep costs down in the expensive cases, hence less generation and storage capacity can be used, reducing asset costs.

5.3.2 Value of building load monitoring

The On-Policy Vol framework presented in Section 2.5.1 was used to quantify the economic benefit of uncertainty reduction from building load monitoring to support the design decision of sizing solar-battery systems and grid connection capacity for the district. A set of 256 hypothetical monitoring measurements were sampled using the probability models defined in Section 5.2.2, and the procedure of using Stochastic Programming to design (size) the system, and simulations to evaluate its performance, was repeated for each hypothesised measurement using samples drawn from the corresponding posterior distribution. This process considers the following questions;

Were the building monitoring systems to have returned a given measurement, reducing uncertainty in the building load profiles: What would the optimal district energy system design be? And what would be its average operational cost under the remaining load uncertainty?

Fig. 5.8 plots the total asset capacities of the optimized system design for each hypothesised load monitoring measurement (the posterior designs, each indicated with a circle), and compares them to the prior system design (indicated by a diamond). It shows that the optimal asset sizes vary significantly with the load monitoring measurements, which provide information about the true energy usage behaviour of the district buildings, by approximately $\pm 20\%$ for each of solar, battery, and grid total capacities. These posterior optimal asset capacities are strongly correlated with the mean load of the posterior distribution, see Fig. 5.9.

A sensitivity analysis studying the influence of input uncertainties on the optimal system design, such as those performed in (Mavromatidis et al., 2018b) & (Pickering and Choudhary, 2019), would stop at this point and conclude that the uncertainty reduction from building monitoring has a substantial impact on the optimal system design, and is therefore important information to gather for achieving a design close to the optimal for the district as built. However, this approach fails to address the actual goal of decision making. From the perspective of the decision-maker, they are not concerned with what decision is taken, but instead with how well on average that decision performs, i.e. its expected benefit/utility. Rather than the difference between the posterior designs and prior design themselves, the metric that is important for determining whether the building monitoring information is valuable for supporting decision making is the difference between their *performance*.

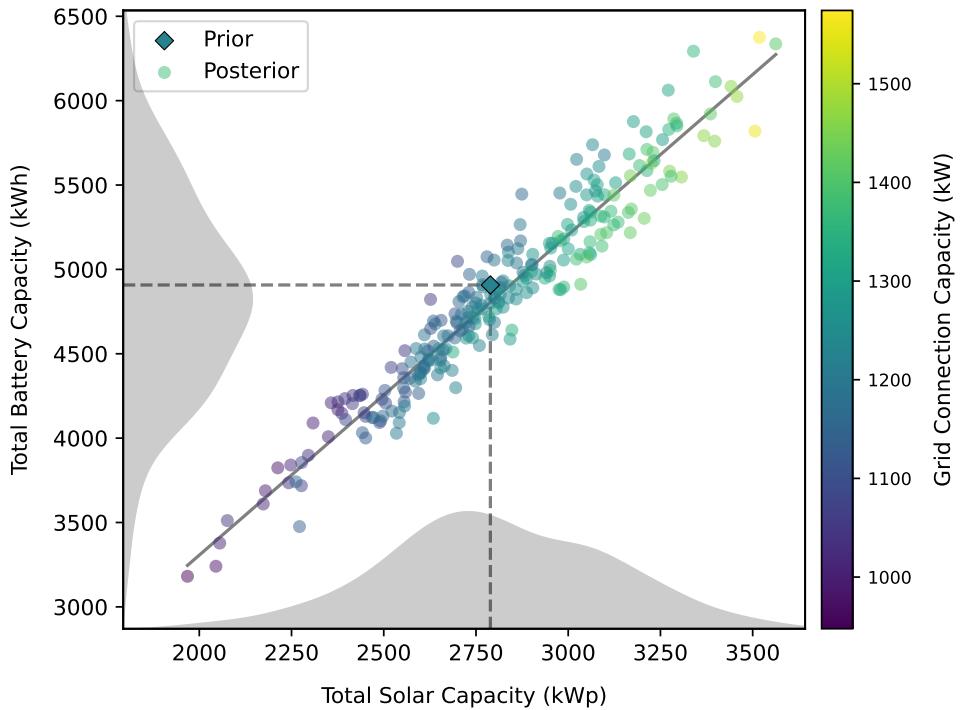


Fig. 5.8 Total asset capacities in optimized system designs for each hypothesised load monitoring measurement. *Prior system design indicated with diamond.*

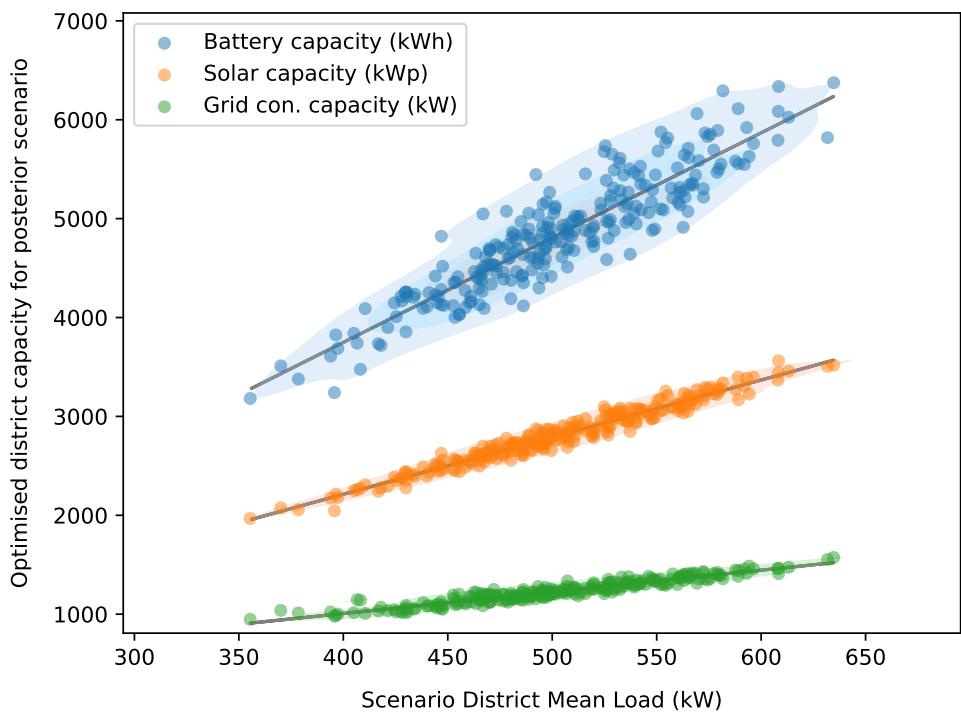


Fig. 5.9 Correlation between total asset capacities in posterior designs and mean load over posterior sampled building load scenarios used for design (from scenario reduction).

i.e. how much lower the lifetime cost of the district energy system is on average as a result of performing the system design with the building monitoring information, and the resulting reduced uncertainty. This metric is called the **Value of Information (VoI)**, defined in Section 2.4.

Figures 5.10 & 5.11 plot the distributions of total system costs and LCOEs achieved by the prior and posterior designs across the simulated building load scenario samples. The average lifetime cost for the posterior designs (the pre-posterior expected utility) is found to be £22.147m over the 20 year lifetime (£1.107m/year). Comparing to the prior expected utility of £22.432m, the **Value of Information** for building load monitoring is estimated at £284k, or 1.27% of the prior cost.

This **VoI** result demonstrates that using building monitoring to support energy system design for this district reduces the average total cost of the energy system by under 1.5%. This methodology leads to a conclusion that is very different from those provided by existing analyses, which suggest that the uncertainty reduction would have a large impact on the design task. Further, the insights provided by **VoI** are far more aligned with the goals of the decision maker.

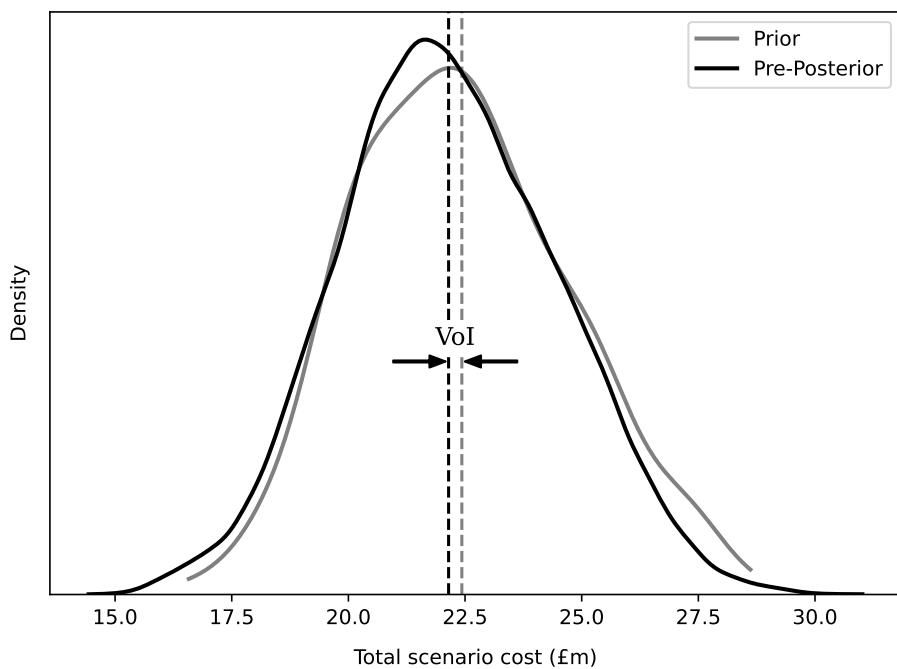


Fig. 5.10 Comparison of simulated total system cost distributions for prior and pre-posterior cases.

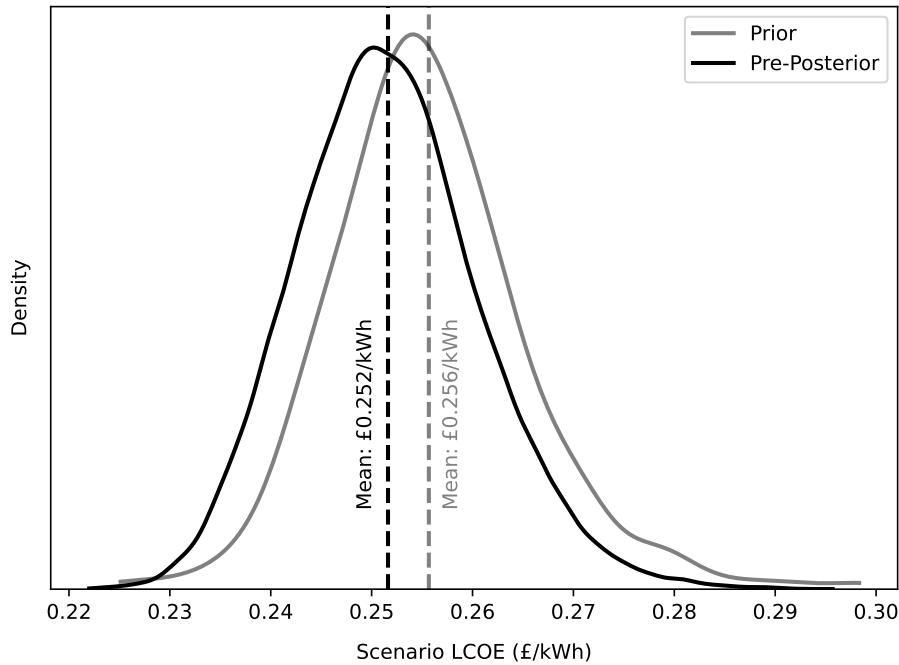


Fig. 5.11 Comparison of simulated LCOE distributions for prior and pre-posterior cases.

Gathering building monitoring data to improve the sizing of the solar-battery systems would only be worthwhile if the cost of collecting that data is less than the **Vol** value computed. Installing the solar-battery systems was found in Sec. 5.3.1 to reduce operational costs by on average £717k/year. Therefore, by itself the option cost of delaying installation of the solar-battery systems for one year to monitor the buildings and install an optimized system outweighs the benefits of that design optimization (provided by the uncertainty reduction from measurement). The actual cost of performing building monitoring would be higher due to costs associated with delaying the construction project. Hence, the **Vol** analysis shows that, for this particular energy system design task, reducing uncertainty in the building load profiles through monitoring is not economically worthwhile. I.e. the benefit it provides to improving decision making is less than its cost.

Relative importance of reduced uncertainties

The relative importance of each aspect of building load uncertainty (type, peak, mean) for supporting decision making was investigated by repeating the **Vol** analysis for uncertainty reduction in each parameter independently. Table 5.3 presents the results of this analysis. Similarly to the classical sensitivity analysis performed in Section 5.3.1, reducing the uncertainty in mean load is found to be of greater importance than uncertainty in the profile shape and peak load, providing roughly half the available benefit to decision making.

Reduced uncertainties	Expected total cost (£m) [%]	Vol (£k) [%]
None (prior)	22.432 [100.0%]	–
type	22.573 [100.6%]	0 [†] [0.0%]
peak	22.401 [99.87%]	30 [0.14%]
mean	22.276 [99.31%]	156 [0.69%]
All	22.147 [98.73%]	284 [1.27%]

Table 5.3 Expected system operating cost and **Vol** for separate uncertainty reduction in load parameters.

While uncertainty in the peak load and load profile shape of the buildings is responsible for a significant portion of the overall cost variability, collecting information which reduces uncertainty in only one of these parameters provides very little improvement for solar-battery system design. In fact, reducing uncertainty in the type parameter, which corresponds to obtaining a probabilistic model of the patterns and timing of energy use specific to each building, does not improve the designs at all. So, by itself, performing detailed occupancy modelling to generate precise time-of-use energy profiles for the buildings would not be helpful for supporting the design of this district energy system.

However, when all three uncertainties are reduced simultaneously, the **Vol** is significantly greater than the sum of the separate contributions⁵⁴. Therefore, if more precise information is available about the mean load of the buildings, then information about the peak load and load profile shape can be used to further improve the design. This result demonstrates that designers must consider not only whether they have the correct information to support their decision making, but the correct combination of information to unlock maximum value. Insights of this kind can only be obtained by applying the **Vol** methodology.

[†]This value is actually estimated to be -£141k (-0.63%), but is clipped to 0 as **Vol** values are defined to be non-negative. This highlights the numerical error in the **Vol** calculations. However, even plus/minus this level of error, the **Vol** values are still small relative to the cost of measurement, and the conclusions remain valid.

⁵⁴This is possible as the **Vol** is not additive. Different measurements can either contain complementary information which is more useful for decision making when combined, or conversely can contain the same information, meaning the combination of measurements provides less total benefit (Di Francesco, Langtry, et al., 2023).

Alternative system scenarios

The same energy system sizing task was investigated in two alternative building system scenarios to study how the **Vol** of building load monitoring varied, and whether the conclusions regarding its economic usefulness changed from the base case. These two cases involve designing for a single building energy system, and designing for a district with constrained solar capacity.

Often adjacent buildings are owned by independent parties, and their energy systems are installed at different times with limited cooperation. So, solar-battery systems are frequently designed independently for each building, and for large consumers (e.g. industrial sites) each must contract their own grid connection capacity. Repeating the **Vol** analysis with a single building, i.e. without district effects, the **Vol** was found to be £99k, or 2.17% of the prior cost of £4.568m. Whilst in this case the **Vol** has increased relative to the prior cost, it is still lower than the annual benefit of installing the solar-battery system, which is £145k/year. Therefore, it is still not economical to perform building monitoring to support design. It is proposed that the reduction in proportional **Vol** as more buildings are included in the district is caused by the variability (standard deviation) of the district mean load. Reducing uncertainty in mean load was found to be the most influential factor for improving system design in Section 5.3.2, and the district mean load was the greatest driver of cost uncertainty, see section 5.3.1. As mean load distributions are taken to be independent for each building, by the Central Limit Theorem, the standard deviation of the district mean load as a fraction of that load decays with the number of buildings in the district as $1/\sqrt{n}$. Therefore, measurement of the buildings provides less information on the district mean load the more buildings there are. Hence, the less valuable that information is for supporting design. This implies the more buildings there are in the district, the better decisions regarding total storage capacity for bulk energy arbitrage average out across the district, and the less sensitive they are to uncertainty in the mean load.

In many European countries, due to land constraints, buildings are built tall to accommodate more floor area, and do not have spare open spaces nearby. This limits their available roof area, and so the quantity of local solar generation they can accommodate to decarbonize their energy usage. The **Vol** analysis was repeated for a district of 5 buildings with a constraint on the maximum solar capacity (kWp) per building. This constraint was derived using roof area data from the Cambridge dataset ([Langtry and Choudhary, 2024](#)). University buildings in Cambridge with around 100 kW mean load ($\pm 20\%$) were found to have a roof area of approximately 1000 m². Assuming a solar panel power density of 0.15 kWp/m² ([Gagnon et al., 2018](#); [Gawley and McKenzie, 2022](#)), leads to an optimistic constraint of 150 kWp of solar capacity per building. With this constraint, the **Vol** of building load monitoring was found to be only £13k, or 0.05% of the prior cost of £29.075m. Under

the formulated operational objective, the primary benefit of the battery system is derived from arbitraging the cheap, zero-carbon solar energy generated by the building solar PV. This is why in the base case, the installed battery capacity of the optimal designs scale linearly with the solar capacity, see Fig. 5.8. For all scenarios in the solar constrained case, the maximum quantity of solar PV is installed, and so the value that can be derived from arbitraging that energy is fixed. Whilst the optimized system designs for each hypothesised measurement in the constrained case show some variability in the battery and grid capacities installed, these additional capacities provide only small additional benefit from arbitraging cheap and/or low carbon grid electricity, and these benefits are not much more than their costs. As a result, updating the system designs (battery sizings) with improved information provides very little benefit in this case.

5.3.3 Planning risk reduction

In the [Vol](#) framework, decision makers are assumed to be risk neutral, and the benefit of data collection is quantified in terms of how much it is able to improve the *average* performance/utility of decisions made. However, in the energy field, decision makers are sensitive to risk, as it affects their ability to raise capital as well as the interest rates they can access, which impacts the cost of energy decarbonization and so their ability to afford it. For the prior optimal design in the base case, the range of system lifetime costs over the simulations was found to be £16.580m to £28.616m, from a mean of £22.432m (see Fig. 5.7a in Section 5.3.1). This 54% cost range could be problematic for the financial planning of the district energy system, particularly as the operational costs of energy system (electricity, carbon, and grid excess costs) vary from £6.360m to £18.396m.

The information provided by load monitoring can be used to support this financial planning, as it reduces uncertainty in the operational cost of energy system. In the experiments, after a hypothetical monitoring measurement was taken, the system design was optimized for the corresponding posterior load distribution, and the operation of this system was simulated for scenarios sampled from that posterior, leading to a distribution of total system costs for the design. Fig. 5.12 plots the distributions over the hypothesised measurements of the standard deviations and ranges of the simulated total system costs for each design (corresponding to a hypothesised measurement), as a proportion of the value for the total cost distribution for the prior design (shown in Fig. 5.7a). The standard deviation and range of the costs are significantly reduced after measurement, demonstrating that whilst designing the energy system with load monitoring information does not significantly reduce the average system cost, at the time of design and so financing, it substantially reduces the uncertainty in what the system cost will be. Therefore, there may be some additional value

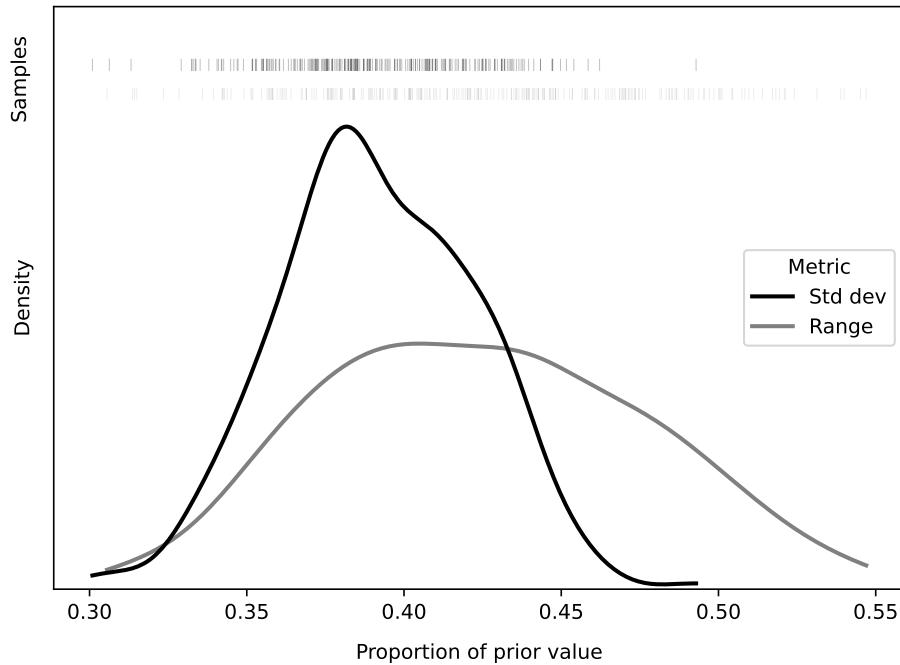


Fig. 5.12 Distribution over hypothesised measurements of standard deviation and range of total cost distribution of optimized system design for each hypothesised measurement (simulated for scenarios drawn from corresponding posterior distribution).

to this information for supporting financial decision making. This motivates the use of [Vol](#) to study the benefit of data collection to support risk averse decision making, which has yet to be done in an engineering context. This would require the expansion of utility/objective functions to account for the cost of risk. Existing studies have incorporated risk measures into district energy system design objectives to account for risk aversion (Pickering, 2019), and these methods are applicable to studying [Vol](#).

5.3.4 Practical importance & limitations of study

The [Value of Information](#) calculations performed in this study (Section 5.3.2) demonstrate that, for the district energy system design task considered, hourly building load data is not worth collecting to support design, i.e. its benefit is less than its cost. This allows practitioners to avoid wasting time and resources on monitoring, and confirms that an opportunity for reducing the cost of decarbonizing building energy is not being missed. Further, showing that limited value is derived from reducing building load profile uncertainty indicates that population level priors constructed from monitoring data from existing buildings are sufficient for designing energy systems in unseen buildings. This provides the first numerical evidence to suggest that current industry practice of designing energy systems

using standard building profiles achieves near optimal design performance. However, it is strongly recommended that using Stochastic Optimization techniques which seek designs that perform well over the distribution of building behaviours is critical.

Section 5.3.2 found that reducing uncertainty in only the mean load provided most of the available decision support benefit. This result is highly pertinent for the design of both new building energy systems, and energy retrofits of existing buildings. For new buildings, while in this case study load monitoring is not worthwhile to support design, the reduction in mean load uncertainty does still provide significant benefit. Therefore, effort should be directed to producing an accurate statistical estimate of building mean load, for instance through detailed building usage modelling, i.e. at room level. Communication between the energy system designer and building owners/occupants regarding the intended use of the building is critical to enabling accurate load estimates, and so low cost energy system designs. For the case of building retrofits, monthly gas and electricity metering data are available. This enables a very good estimate of the mean building electrical load after retrofit, and so using hourly metering data for system sizing would provide little additional benefit.

In this study, uncertainty in building load and solar generation only were considered for the sake of clarity when demonstrating the [Vol](#) methodology. However, several other important factors in the district energy system are subject to significant uncertainty, such as electricity prices (which are determined by the future grid generation mix and weather conditions), asset operational lifetimes which impact their effective cost, and battery degradation (neglected in this model) which affects the efficiency and capacity of the battery units over their lifetime. More detailed system design work should consider these uncertainties in the Stochastic Optimization model to include their influence on the design task and resulting cost distributions. A key difference with these uncertainties is that they are not practically measurable at the time of design, and so should not be studied using [Vol](#), as the resulting values would not be meaningful.

Performing [Vol](#) calculations is highly computationally intensive. In the On-Policy [Vol](#) framework, system designs must be both determined and evaluated for a large number of hypothesised measurements to achieve sufficient statistical accuracy in the mean utility estimates to enable the correct identification of the true [Vol](#) value over statistical error. This limits the complexity/resolution of building energy system models that can be studied with the framework. However this computational cost is minimal compared to the potential cost savings of avoiding project delays to gather unnecessary building monitoring data. The computational requirements of the [Vol](#) calculations from this study are discussed further in 5.C.

Finally, the key critique of the **Vol** framework is that **Vol** values computed are only valid under the particular model formulation studied, including its assumptions (Langtry, Zhuang, et al., 2024). **Vol** results cannot be generalized, and changes in the problem setup can significantly alter the **Vol**, as seen when introducing solar capacity constraints in Section 5.3.2. Therefore, further investigation of **Vol** for supporting building energy system design is required to confirm whether the conclusions made in this study are consistent across different building energy systems (e.g. those in different climates or with different energy usage patterns), design objectives, and modelling assumptions. The **Vol** methodology should also be used to investigate the importance of reducing other uncertainties in buildings (such as thermal properties of the envelope), and whether performing other types of measurements could provide significant improvements for the design of energy systems.

5.4 Conclusions

This chapter investigated the economic benefit of gathering hourly building energy usage data to support the design of a district energy system, specifically involving the sizing of distributed solar-battery systems to decarbonise building energy usage under grid constraints. A case study district energy system containing 5 university buildings was considered, with probabilistic models of the building load profiles constructed using historic electricity metering data from the Cambridge University estate.

The uncertainty in building load was found to have a significant impact on the operating cost of the system, causing a $\pm 30\%$ variation in total cost for the energy system designed without monitoring data available. Also, reducing uncertainty in the building load by collecting monitoring data had a significant effect on the optimized energy system design, with total installed capacities of solar, battery, and grid connection varying by roughly $\pm 20\%$ across the hypothesised measurement values. Therefore, existing methods which stop at either of these two points would conclude that collecting hourly metering data would have a significant effect on system design.

However, performing a **Value of Information** analysis using the On-Policy **Vol** framework demonstrated that reducing load uncertainty via monitoring would reduce the overall energy system cost by less than 1.5% on average through improved design. This was found to be less than the cost savings achieved by installing solar-battery systems to reduce operating costs over the period needed to monitor the buildings and install systems with improved design. Therefore, collecting hourly building monitoring data was determined to be not economically worthwhile for supporting the design of the studied district energy system. Reducing uncertainty in only the mean load was found to provide most of the available

decision support benefit. This means that for building retrofits where the mean building load can be well estimated from historic monthly energy usage data, using hourly metering data would provide little additional benefit for system design. So while Chapter 4 found that 1 to 2 years of load monitoring data is needed to train a high accuracy hourly prediction model of a building's electricity usage, it is not necessary to have this data at the time of designing the solar-battery systems (district energy system). This means that when the battery systems are initially installed, they must use controllers trained on data from other buildings, as no hourly load data from the new building would be available. This can either be done by re-using a model from an existing building, which was investigated in Section 4.3.2, or by training a generalised/foundation model for the population of buildings, which is explored in (Raisch, Langtry, et al., 2025). These results demonstrate the need for **Value of Information** analysis to justify and prioritise expenditure on data collection to support decision making, as current analyses can lead to erroneous conclusions regarding the benefit of uncertainty reduction.

These results are important for practitioners, as identifying the low value of load uncertainty reduction for supporting district energy system design allows wasteful delays to monitor building loads⁵⁵ to be avoided. They also indicate that population level priors constructed using monitoring data from existing buildings are sufficient for energy system design. This provides the first numerical evidence to support the sufficiency of existing industry practice of using *distributions* of standard building load profiles⁵⁶ for energy system design.

However, as the **Vol** framework provides no generalisation guarantees, these results can only indicate that collecting load monitoring data is not worthwhile for supporting the design of the particular case study district investigated. Further research is required to determine whether the benefit of load monitoring remains below its cost across different energy system setups, different design objectives, and with the inclusion of further uncertainties. Additionally, the load monitoring data may provide value to the financial planning of the district by reducing uncertainty/risk in the total system cost. In the next chapter, the value of uncertainty reduction for reducing risk in decision making is investigated by introducing risk aversion to the decision objective.

⁵⁵ Note that monitoring systems would still be installed in the buildings, as real-time building load data is critical for effectively controlling the solar-battery systems. This work quantifies the value of having access to this data *before* the solar-battery systems are designed, to reduce uncertainty in the building loads and improve sizing decisions. The results demonstrate that it is not worthwhile delaying the installation of the solar-battery systems to collect data from the monitoring systems, as the benefit to decision making is less than the cost of doing so. Over time, operational data from the monitoring systems can be used to improve the building load distributions which are then used for future designs.

⁵⁶ Many building designers still use deterministic building load profiles (called design years) for energy system sizing. As discussed in 2.E.1, these do not necessarily provide a good representation of the possible building behaviour in the design optimisation. The results of this chapter only support the sufficiency of using load profile distributions from the population of existing buildings in a *stochastic* design optimisation. Further research is needed to investigate how much value is being lost from the use of deterministic design years in building energy system design, i.e. the **VSS** (see Section 2.2). The code and data from this chapter could be easily adapted to do this analysis.

Appendix 5.A Case study system parameter specification

Common parameter values specifying the district energy system, probabilistic model of building load, and procedure for sampling from that model, are used across all experiments. Table 5.4 provides values for the parameters of the district energy system model, including costs defining the objective/utility, simulation durations, and assumptions regarding battery performance and grid control. All parameters for the probabilistic model of building load are specified within the definition of the probabilistic model in Section 5.2.2. Table 5.5 provides the settings used when sampling from the defined distributions, and performing scenario reduction using the Fast Forward technique (Heitsch and Römisch, 2003), as implemented in (Gioia, 2023).

Parameter	Units	Value	Note/Refs
Carbon cost	£/kgCO ₂	1	Chosen to be much greater than current carbon trading prices, c. £0.1/kgCO ₂ (Bloomberg, 2024; DESNZ, 2022) to reflect building owners' decarbonization priorities.
Battery capacity cost	£/kWh	750	(Forbes Home, 2024; Mott MacDonald, 2018)
Solar capacity cost	£/kWp	1500	(Gawley and McKenzie, 2022), (E.ON Energy, 2024); corresponds to LCOE of £58.6/MWh, in line with expected value (DESNZ, 2023b)
Grid connection capacity cost	£/kW/day	0.263	Chosen to be much greater than current costs, c. £0.05/kVA/day (Networks, 2023), to reflect future grid constraints. Converted to per kW value assuming a power factor of 0.95 (EEPower, 2022), (Anderson et al., 2021).
Grid connection excess charge	£/kW/day	1.053	As above. Current value c. £0.08/kVA/day (Networks, 2023).
Simulation duration, T	Hours	8760	
System lifetime	Years	20	
Design grid safety factor	–	1.25	Determined from initial experiments.
Operation grid safety factor	–	1.01	Determined from initial experiments.
Battery discharge ratio, δ	kW/kWh	0.4	Value for Tesla Powerwall (Forbes Home, 2024)
Initial state-of-charge, SoC ⁰	–	0	

Table 5.4 Parameter values for district energy system model used in experiments.

Parameter	Value
No. samples from prior distribution	1000
No. samples from each posterior distribution	256
MCMC sampling burn-in period	256
MCMC sampling thinning factor	10
No. of reduced scenarios used in Stochastic Program	10

Table 5.5 Settings for sampling from probabilistic load model used in experiments.

Appendix 5.B Details of prior system design

The sizing of grid connection capacity and solar-battery systems for each building determined by the prior design optimization is presented in Table 5.6. As all buildings in the district are taken to have identical probabilistic load models and solar generation potentials, in the perfect case, the design optimization would allocated equal asset capacities in all buildings, as on average they behave identically. However, Table 5.6 shows significant variation in the solar-battery sizings across the buildings. This occurs as only a small number of reduced scenarios (10) are included in the Stochastic Program model, due to the cubic computational complexity of Linear Programming limiting the number of scenarios which can be considered. Random variation in the mean building loads across the reduced scenarios results in non-uniform sizing across the district.

Table 5.7 provides a breakdown of the system operating cost contributions for the prior design across the 256 simulated building load scenarios.

Building no.	Solar capacity (kWp)	Battery capacity (kWh)	Grid con. capacity (kW)
0	617	1,119	—
1	538	973	—
2	497	844	—
3	583	983	—
4	554	989	—
Total	2,789	4,908	1,227

Table 5.6 Installed asset capacities for prior system design.

Cost	Mean (£m)	Percentage (%)	Std. dev. (£m)
Total	22.432	100.0	2.282
LCOE (£/kWh)	0.256	—	0.00971
Electricity	7.416	33.1	1.357
Carbon	4.795	21.4	0.934
Grid excess	0.0	0.0	0.0
Grid connection	2.356	10.5	—
Battery	3.681	16.4	—
Solar	4.183	18.6	—

Table 5.7 Breakdown of cost contributions over sampled scenario simulations for prior system design.

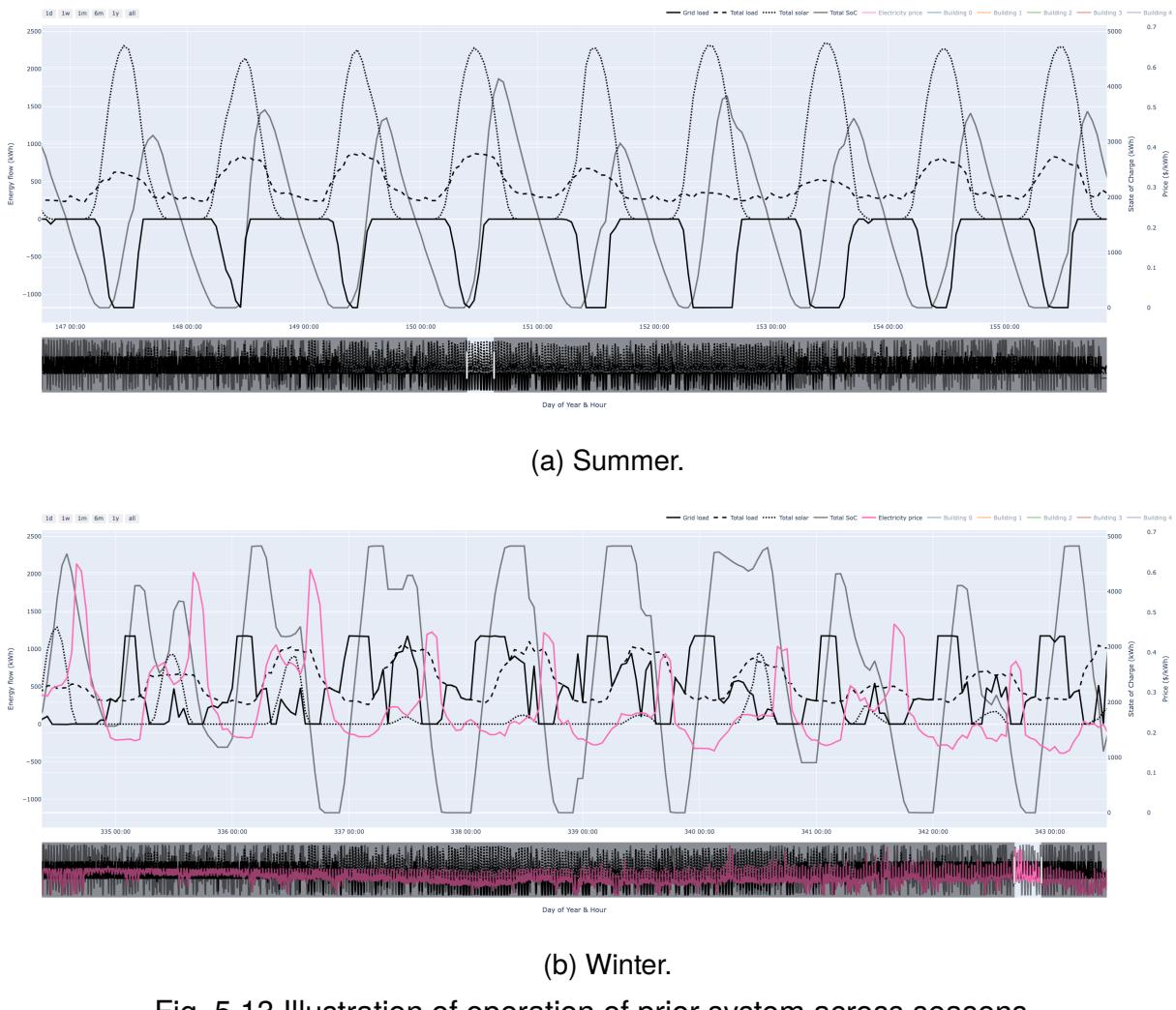


Fig. 5.13 Illustration of operation of prior system across seasons.

Interactive version of plot available online [here](#).

Observation of the operational behaviour of the district energy system provides insight into the factors driving system design. Fig. 5.13 plots the overall loads and battery state-of-charge in the district when operating over periods in both the summer (Fig. 5.13a) and the winter (Fig. 5.13b) for one of the sampled load scenarios. Readers can investigate the operational behaviour of the district energy system further using an interactive version of plot available online at https://mal84emma.github.io/Building-Design-Vol/example_prior_system_simulation.html.

Comparing Figs 5.13a & 5.13b shows that the net peak powers caused by local solar generation in the summer are significantly larger than those caused by building load in the winter. As a result, in the base case without constraints on the solar capacity, the required grid connection capacity is determined by the quantity of solar PV, and the optimization trades-off grid connection cost with additional cheap, zero-carbon solar energy. This implies that consideration of uncertainty in solar generation in the design optimization is important,

as the capacity factor of solar generation will determine the scaling of this trade-off (noting that the peak power is well known as it is approximately the capacity in all years). This behaviour of the district energy system regularly exporting large quantities of energy onto the grid at around noon in the summer may cause issues for regulation of the power grid, as discussed in (Iweh et al., 2021), and will require either adaptation of grid infrastructure to handle reverse power flow, or differentiated pricing for grid export capacity to discourage energy prosumers (customers with local generation) from flooding the grid with solar energy at these peak times. This may then increase the need for local storage to prevent surges in grid export to manage grid stability.

For the majority of the summer, the district draws no power from the grid, though it provides significant quantities of energy to the grid. Therefore, over the summer this district would have zero energy bills. Additionally, as in this model the energy system operator is not paid for supplying energy to the grid, the solar generation capacity is therefore being set by its ability to provide cheap, zero-carbon energy during the spring and autumn to offset grid electricity. It is therefore correlated with the district mean load, as this determines the quantity of energy which must be offset. In future energy systems with large quantities of solar generation, it is likely that during peak generation times the price of grid electricity would fall to approximately zero, or potentially negative, as instantaneous generation exceeds energy demand.

Full depth battery charge cycles are observed in the winter but not in the summer. This indicates that the battery capacities are determined by the financial benefit of arbitraging cheap grid electricity for use by the buildings at peak times. These capacities are then sufficient for maximising the usage of local solar energy generation, and managing power peaks to avoid exceeding the grid connection capacity. Fig. 5.13b shows that the Model Predictive Controller successfully performs this price arbitrage of energy. On most days of the simulation, the power draw from the grid is near zero at peak demand times (around 5pm). In fact, this distributed generation and storage system reverses the timing of the peaks as the controller is not directed to smooth power usage.

As the district mean load determines the quantity of energy which must be arbitrated to reduce the cost of satisfying building loads at peak times, via the same mechanism as for solar capacity, the optimal battery capacity is therefore also determined by the district mean load. This justifies the correlation between optimized solar and battery capacities observed in Fig. 5.8.

Appendix 5.C Investigation of numerical accuracy & computational requirements

5.C.1 Statistical accuracy

As the **Vol** is the difference between two expectations, the prior and pre-posterior expected utilities (see Eq. 2.14), which are estimated using Monte Carlo sampling, the computed **Vol** values are statistical estimates of some underlying true value. The statistical accuracy of these estimates must therefore be considered before firm conclusions can be drawn regarding whether data collection is worthwhile.

Fig. 5.14 plots the convergence of the prior and pre-posterior expected utility estimates, as well as the **Vol** estimate, as the number of samples from the building load distribution (scenarios, with corresponding designs and simulations) used increases. It also plots 95% confidence intervals for the expectations using the standard error on the sample mean. For the **Vol** estimate, the **Vol** is considered as a single expectation, and the standard error is computed for the convergent sequence of **Vol** estimates. The expectation estimates are seen to converge within the 256 samples used in the experiments, providing very similar estimates for all metrics after only 128 samples. This indicates that the **Vol** values computed in the experiments have sufficient statistical accuracy to support the hypothesis that the **Vol** is smaller than the cost of measurement, and therefore building monitoring is not financially worthwhile for supporting system design.

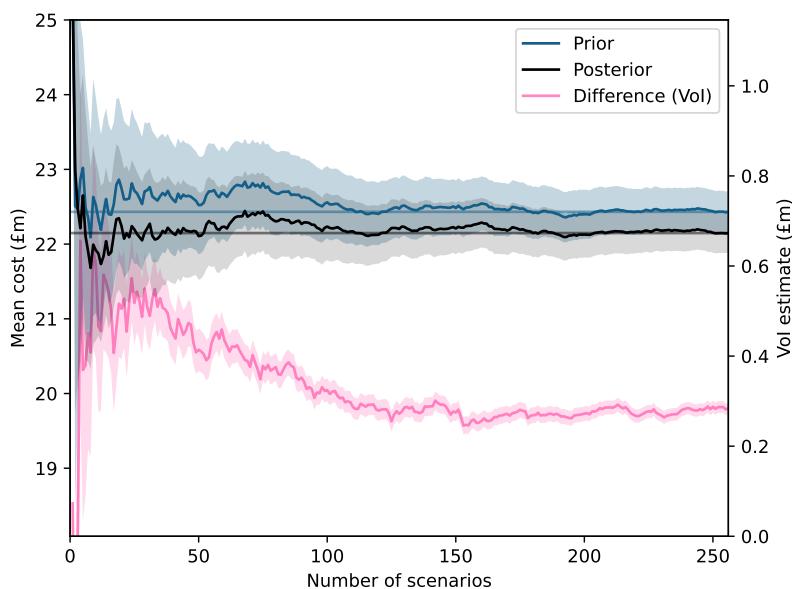


Fig. 5.14 Convergence of Monte Carlo estimates of expected utilities with number of samples from building load distribution.

In this study a naive sampling scheme was used to sample from the building load distribution. The use of improved sampling techniques which provide greater sample efficiency for the load distribution would speed up the convergence of the expectation estimates, and so reduce the computational cost of achieving sufficient statistical accuracy in the **Vol** estimates. Further work is required to properly quantify the statistical error of **Vol** estimates, and develop techniques for determining the number of samples required to achieve sufficient statistical accuracy to support conclusions drawn from the **Vol**, i.e. whether it is greater or less than some cost threshold. Such developments are necessary to reduce the computational cost of **Vol** calculations, to allow the study of complex, practical engineering decision problems, whilst providing confidence that the data collection strategy recommendations they produce are correct with high probability.

5.C.2 Cost estimation accuracy

The On-Policy **Vol** framework proposed in Section 2.5.1 enables the **Vol** to be computed for complex decision problems where policies which do not necessarily provide an accurate estimate of the decision utility must be used for computational tractability. For the case study considered in this work, Stochastic Programming (SP) is needed to optimize the energy system sizing as the number of decision variables and simulation time make the application of stochastic global optimization methods directly to the simulator intractable. However, the SP's estimate of mean system cost (expected utility) suffers from two sources of inaccuracy.

Firstly, the SP can only include 10 scenarios in its Monte Carlo estimate of mean operational cost within its objective function to run in a manageable computation time. This leads to potentially significant statistical error in the estimate of cost that is minimized. An attempt is made to manage this by using scenario reduction to select the 10 samples/scenarios that provide the best statistical representation of the population. However, as the space of possible load profiles is extremely large, only simple summary statistics could be used as the targeted statistical metric to represent, meaning the selection of scenarios to include in the SP does not have a direct link to the distribution of operational costs. Secondly, modelling error is introduced as the Linear Program formulation of the decision problem requires the assumption that the energy system is controlled with perfect foresight of the operational conditions, i.e. the weather and electricity price are known perfectly for all future periods during control. This assumption is not realistic and can lead to the SP model underestimating operational costs as the system is able to pre-empt difficult operational conditions and make unrealistically good control decisions in advance to reduce overall cost. In the case of managing the grid impact of energy usage under load uncertainty this assumption can be particularly over-optimistic, as unforeseen load conditions can lead

to the battery state-of-charge not being sufficient to manage large peak loads, leading to substantial grid excess charges. Assuming perfect foresight allows the system to manage the battery level to avoid these instances, and causes a significant underestimate of operational cost. In fact, it is this over-optimism regarding the system's ability to manage peak loads that leads to the need to include a factor-of-safety on grid capacity during design.

Therefore, as these model and statistical errors could lead to significant inaccuracy in the SP's estimate of operating costs, a simulation step was included to provide a more accurate estimate of the true costs. These simulations used a Model Predictive Controller with only 48 hours of perfect foresight⁵⁷, and evaluated a large number of sampled load scenarios to overcome both the model and statistical inaccuracies.

The accuracy of the Stochastic Program model objective was investigated by comparing it to the mean system operating costs determined by simulation. Table 5.8 provides both the mean error and mean absolute error of the SP objective for system designs performed with both the prior and posterior distributions, across the range of cases considered in the experiments. Fig. 5.15 plots the distribution of objective errors for the SP designs performed for each measurement scenario in the base case, where all uncertainties were reduced.

Case	Mean error (%)	Mean absolute error (%)
Prior cases (as considered in Section 5.3.2) - point estimates		
Base	-0.345	0.345
Solar constrained	-0.423	0.423
1 building	-3.062	3.062
Posterior cases (as defined in Table 5.3 of Section 5.3.2) - av. over 256 msrmnts.		
type	0.683	0.811
peak	-0.113	0.899
mean	-0.633	1.052
All (base)	-0.271	0.525

Table 5.8 Errors in SP estimates of mean system operating costs for optimized designs (under-estimation +ve).

The results show that the Stochastic Program is able to estimate the mean operating cost within around 0.25-1% (except for the single building case where the error is significantly higher), and in fact more frequently over-estimates the cost than under-estimates. This

⁵⁷This controller is still unrealistic, as perfect foresight even over a limited planning horizon leads to practically unachievable low operational costs, as demonstrated in (Langtry, Wichitwechkarn, et al., 2024). While it would not be feasible to train practical prediction algorithms to use in each simulation, a controller using synthetically noisy forecasts calibrated to the properties of practical forecasting methods could be used to improve the realism of the simulated operation and resulting costs.

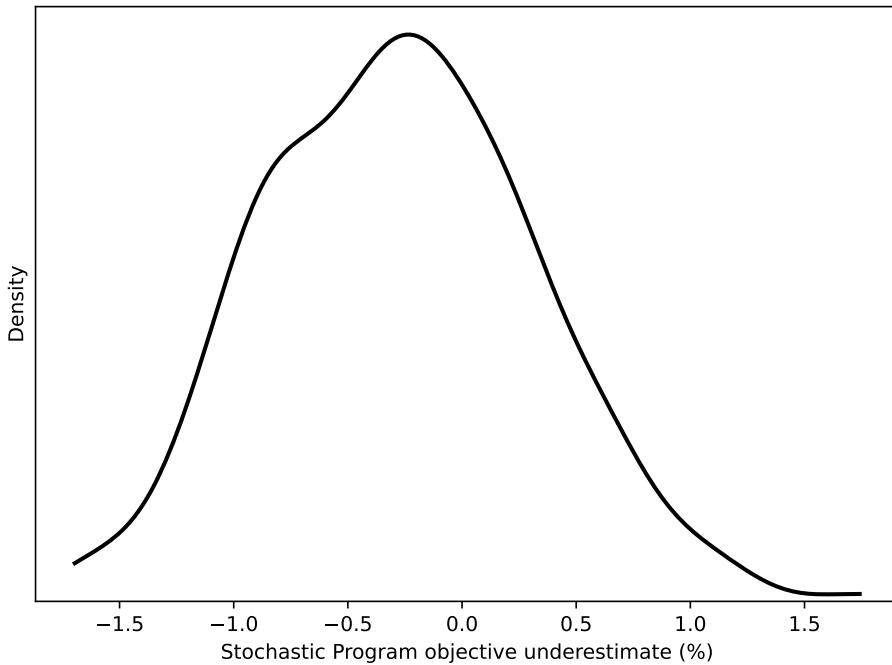


Fig. 5.15 Distribution of Stochastic Program objective value errors compared to simulated mean operating costs for posterior designs in base case.

result is somewhat surprising considering that only 10 scenarios are considered in the SP objective, and these scenarios were selected in a simplistic manner. From the plots of operational behaviour (Fig. 5.13) it can be seen that battery charge is rarely held for more than 72 hours, which indicates that during operation, assumption of complete perfect foresight would not lead to significant under-estimation of operational costs. However, this also demonstrates that the factor-of-safety used during design is performing well as the system does not end up in situations where it does not have the state-of-charge needed to manage peak loads. Additionally, as district mean load was one of the metrics used in scenario reduction, and was found to be the key driver of operational cost (Section 5.3.1), the scenario reduction was able to well represent the distribution of costs through correlation.

However, whilst the SP objective error is relatively low, it is approximately the same magnitude as the **Vol**, which was 1.27% in the base case. An additional 1% cost error added to the **Vol** estimate could have led to the conclusion that building monitoring was economically worthwhile, as it would have put it at around the same level as the estimated cost of performing the measurement. However, the impact of the SP objective error on the **Vol** estimate is complex, as it depends on the error in the difference between the prior and pre-posterior cost estimates, which is challenging to study. Ultimately, verifying that the error is sufficiently low may be as computationally expensive as performing the simulations required by the On-Policy **Vol** calculation.

5.C.3 Computational cost

The computational cost of the experiments performed in this study was very large, with each **Vol** calculation taking roughly 10 days of computation time on a powerful desktop workstation⁵⁸ with a parallel processing workflow. Performing the simulations took approximately 80% of runtime.

Reducing the number of samples required by the Monte Carlo estimates of expected utilities, or removing the need for simulation to accurately estimate utilities, would greatly decrease the computational cost of **Vol** calculations. Given that the SP objectives are found to be reasonably accurate estimates of the simulator costs in 5.C.2, a good estimate of the **Vol** could be obtained in only 20% of the original runtime. A key advantage of **Vol** calculations is that they are highly parallelizable, as each scenario can be evaluated independently, making them well suited for distributed computation. I.e. the optimization based design and simulator evaluation can be performed simultaneously for each measurement scenario, meaning the computation time can be drastically reduced if sufficient compute resources are available.

⁵⁸ 32 core, 2.90 GHz Intel Xeon Gold CPU with 256 GB of DDR4 RAM.

6

Hedging storage technology selection for energy park design

Max Langtry and Ruchi Choudhary (Oct. 2025b). “The Value of Hedging against Energy Storage Uncertainties When Designing Energy Parks”. In: *Energy* 334, p. 137600. ISSN: 0360-5442. doi: [10.1016/j.energy.2025.137600](https://doi.org/10.1016/j.energy.2025.137600). URL: <https://www.sciencedirect.com/science/article/pii/S0360544225032426>

This chapter has been published as the journal article above.

All code and data used to perform the experiments in this chapter is available at <https://github.com/mal84emma/Energy-Park-Design>.

Grid scale energy storage needs to be built quickly and in large quantities to support variable renewable generation. Co-locating generation and storage with industrial loads, referred to as energy parks, allows the renewable supply to be matched the energy demand, enabling greater decarbonisation while also reducing operating costs and grid impact. However the performance of bulk energy storage technologies is currently highly uncertain. As technology development progresses and uncertainty in storage performance reduces, a different technology may turn out cheaper than the one chosen at the start of the energy park project. Enabling flexibility in the design process so that system designs can be updated as better information becomes available would lower the cost of decarbonising industry. But achieving this flexibility is itself costly. This raises the question, “Is it worth it?”. This chapter quantifies the benefit of retaining flexibility to adapt energy park designs and optionality over storage technology choice as uncertainty reduces, to determine whether they are economically worthwhile. It applies the [Value of Optionality \(VoO\)](#) framework from [Section 2.5.2](#) to the sizing of wind, solar, and storage in an illustrative energy park model based on a real-world proposal near Rotterdam, considering uncertainty in storage efficiency, lifetime, and capital cost. The chapter investigates the reduction in the total cost of the energy park when the asset sizings and energy storage technology can be changed after uncertainty in storage performance has been reduced by building demonstrator storage systems. It also considers how this changes when multiple storage technologies are used in the energy park, and when the energy park developer is risk averse.

6.1 Introduction

Currently 40% of electricity in the UK is generated from carbon intensive sources (DESNZ, 2024a), and electricity demand is expected to increase by a factor 3 to 4 as heating and transport energy usage are decarbonized via electrification (ESO, 2023). As a result, around 100 GW of wind and 60 GW of solar generation need to be constructed by 2050 (ESO, 2023). However, managing the variability of electricity generation from these renewable sources to maintain security of supply presents a significant challenge (Papadis and Tsatsaronis, 2020), and the required adaptation of the energy system will be costly. Grid-scale energy storage is expected to be the main mechanism for matching renewable electricity supply with demand. Co-locating bulk energy storage with renewable generation and industrial loads, referred to as an energy park, has been proposed as a way of both improving the profitability of grid-scale storage (reducing the cost of its services), and reducing the grid impact of variable renewable generation (Chinaris et al., 2025; F. Fan et al., 2021).

The design of energy parks, including the selection and sizing of renewable generation and energy storage technologies, has been extensively studied in the literature. For example, Arévalo & Jurado (Arévalo and Jurado, 2021) investigates the combination of energy storage technology and control scheme that provides the lowest net present cost for supporting an autonomous grid with solar, wind, hydro, and diesel generation. Sizing optimization is performed for each system configuration, and it is found that due to the different performance characteristics of the storage technologies, they are best suited to supporting substantially different generation mixes. This indicates that the technical performance of energy storage has a significant impact on how the energy system should be sized. Hu et al. (Y. Hu et al., 2024) considers a district energy system with local renewable generation providing both electricity and heat. It demonstrates that optimizing the sizing and location of battery storage, hydrogen storage, and CHP units in the system can reduce the net load fluctuation by 25%. Similar benefits of using energy storage to manage variability in both renewable generation and load are shown by Phu et al. (Phu et al., 2024), which studies an energy park producing green hydrogen from biomass using electricity from wind and solar generation. Optimizing battery and hydrogen energy storage within the system is shown to substantially reduce the operating cost, carbon emissions, and grid reliance of the energy park. Further this optimized configuration is demonstrated to be robust against uncertainties in hydrogen demand, grid electricity price, and renewable generation.

Many different energy storage technologies have been considered for supporting energy parks, such as Lead-acid batteries, Lithium-ion batteries, compressed-air storage, and redox flow batteries (IRENA, 2017). However the future performance of these technologies,

i.e. how they will perform when implemented in the grid from around 2030 onwards, is still highly uncertain. Across the literature, estimates of technical parameters (e.g. round-trip efficiency and self-discharge rate) and economic parameters (e.g. cost per energy capacity and service lifetime) vary over wide ranges (IRENA, 2017; Kebede et al., 2022; Kittner et al., 2020; NESO, 2022; Petkov and Gabrielli, 2020). For example, a literature review, (Kebede et al., 2022), found estimates for Sodium–Sulphur high-temperature batteries ranging from 70–90% for round-trip efficiency, and 0–20% for self-discharge rate. (Petkov and Gabrielli, 2020) found values for the capital cost of Li-ion batteries between 150 and 600 €/kWh, and operational costs between 1 and 5%. And an IRENA report, (IRENA, 2017), estimated the annualized energy capacity cost of compressed-air storage to be between 2 and 355 cents/kWh/yr.

These highly uncertain technical and economic performance parameters have a significant impact on the cost of developing and operating grid-scale storage in energy parks. Crucially, it also causes uncertainty in which storage technology will provide the lowest cost pathway to supporting variable renewable generation. As research and development of storage technologies progresses, uncertainty in their performance will decrease, and the best technology will emerge. However, to meet 2030 and 2040 carbon emissions reductions targets, energy parks must be designed and begin development in the near future.

This raises the following question, “When designing an energy park, can the best energy storage technology be identified with the current level of uncertainty in performance? Or should developers keep their options open to potentially make a better choice when uncertainty has reduced?” No existing studies have considered the impact of reduction in storage performance uncertainty on the design of energy parks and the selection of storage technologies. In fact, very few consider storage uncertainties during energy park design.

6.1.1 Storage technology performance uncertainty in energy park design

The impact of uncertainty on the operation of energy parks, and so the cost of arbitraging energy to match variable generation with demand, has been widely studied (Ji et al., 2025; J. Kim et al., 2017; P. Li et al., 2021; T. Zhang and Sobhani, 2023). For instance, (J. Kim et al., 2017) shows that accounting for load forecast uncertainty in control can reduce operating costs by 10%, and (Ji et al., 2025) develops a method for managing the risk introduced by uncertainty in load and solar generation during operation. Many previous articles have investigated the optimal sizing of generation and storage assets to maximize total profit in the presence of various uncertainties. For example, Kim et al. (S. Kim et al., 2024) consider

the sizing of wind generation, battery storage, and electrolyser capacity in a micro-grid with local demands for both electricity and hydrogen. A bi-level optimization method is used to minimize the average cost of meeting the local energy demands, accounting for uncertainty in wind generation and demand patterns on operating costs, and uncertainty in the per capacity costs of the assets. Gabrielli et al. (Gabrielli et al., 2019) develop a method for sizing a multi-energy system with solar generation, battery storage, heat pumps, and an electrolyser, so that the design is robust to uncertainty in weather conditions and the electricity and heating demands which must be met. Bakke at al. (Bakke et al., 2016) study the optimal level and timing of investment in stand-alone battery storage for providing energy arbitrage and ancillary grid services, considering uncertainty in both electricity prices (which determine operational revenue) and evolution in the price per capacity of battery storage.

Uncertainty in renewable generation (Gabrielli et al., 2019; S. Kim et al., 2024), energy demands (Keyvandarian and Saif, 2023; Maia et al., 2021), and the price of grid electricity (Bakke et al., 2016; Mohammadi et al., 2022) and carbon (Yongbiao Yang et al., 2024) are commonly considered during energy park design. However, few studies account for uncertainties related to storage technologies. Some have included uncertainty in the cost of energy storage (Bakke et al., 2016; Chadly et al., 2023; Coppitters et al., 2021; S. Kim et al., 2024). However only two studies could be found that consider uncertainty in the technical performance of storage technologies, (Coppitters et al., 2021) which accounted for uncertain cycle life and self-discharge rate, and (Chadly et al., 2023) which considered uncertain round-trip efficiency.

No existing works have investigated the impact that uncertainty reduction has on the process of energy park design. Additionally, in the broader energy systems literature, risk aversion is known to be an important feature of decision making (Mu et al., 2022; Pickering, 2019; Tostado-Véliz et al., 2024). Its effect on the selection and sizing of energy infrastructure in parks has been investigated in studies such as (Mu et al., 2022) and (Tostado-Véliz et al., 2024) respectively. However, the impact risk aversion has on the benefit of uncertainty reduction is yet to be studied⁵⁹.

⁵⁹ Some early work investigated the effect of risk on *Vol* (Gould, 1974). However this explored only the theory and simple demonstration problems. No studies have investigated the effect that risk preferences have on *Vol* in complex, real-world decision making.

6.1.2 Storage technology selection in energy park design

While the majority of energy park design studies consider only a single storage option, a few do compare multiple storage technologies to identify the option best suited to the energy system. Zhang et al. (Y. Zhang et al., 2019) compares the use of hydrogen storage and Lithium-ion batteries to manage the variability of wind generation and reduce curtailment. (Mazzoni et al., 2019) & (Mohseni and Brent, 2022) compare different battery storage technologies for supporting distributed energy systems with local renewable generation. They find that Lead-acid and Sodium–Sulphur batteries provide the lowest total operating cost for their respective energy systems. However, none of these studies account for uncertainty in the characteristics of the storage technologies during the design optimizations which are then compared. A comparison of battery technologies which does account for performance uncertainties is performed in the context of building energy systems in (Chadly et al., 2023).

The main limitation of all these existing studies is that they consider the energy park design process, including both technology selection and sizing, as a static, single-point decision, where the entire design is determined up-front. This does not reflect the design process of practical energy systems, where several stages of engineering design are performed progressively with increasing detail, and greater information regarding available components (which for instance may be provided during contracting).

As uncertainty in the performance of the available storage technology options reduces, the best technology choice (that with, for example, the lowest average cost under the remaining uncertainty) may change. As a result, retaining optionality over which storage technology is ultimately used through the design process could be valuable, as it may be possible to reduce the overall cost of the energy system by selecting a better performing technology at a later stage. However, keeping this optionality and gathering improved information about the performance of each storage technology is costly. This raises the question of whether the benefits of retaining storage optionality are worth the cost.

6.1.3 Research objectives & contributions

There has been no study of the impact that reducing uncertainty in energy storage technology performance has on the design of energy parks or the selection of storage technologies. As there is currently large uncertainty in the performance of prominent storage technologies, it is important to understand whether developers are able to commit to a single technology up-front, or whether it is beneficial to hedge their bets and have the option to choose the best technology when better information about performance is available.

This chapter uses the [Value of Optionality \(VoO\)](#) framework to numerically answer the following questions, “How valuable is retaining optionality in storage technology choice for the design of an energy park as uncertainty reduces? Is this benefit to design worth the cost of obtaining improved estimates of storage performance and retaining the option to change storage technology?”. It investigates this in the context of sizing wind, solar, and storage capacities for an illustrative energy park system, modelled off a real-world green hydrogen plant proposal in the Port of Rotterdam. The main research objectives are to:

- Quantify the benefit of updating the system design (sizings) after uncertainty in storage performance has been reduced
- Determine whether the benefit to design of retaining optionality in storage technology selection by speculatively developing multiple technologies is worth its cost
- Investigate how the level of uncertainty reduction and risk aversion impact whether design adaptation and storage optionality are worthwhile

The key research contribution is that this work is the first to investigate the impact of reducing uncertainty in storage technology performance on energy system design, and to quantify the benefit of retaining optionality in storage technology selection as uncertainty reduces. The analysis in this chapter demonstrates the value to energy system developers of providing the opportunity to update system designs as improved information about the performance of energy storage (with reduced uncertainty) becomes available, which has important implications for industry design practice.

The remainder of this chapter is structured as follows. Section 6.2 describes the illustrative energy park studied, including the probabilistic models of renewable generation, industrial load, and energy storage technology performance uncertainty used, and the Stochastic Programming model used to perform system design. Section 6.3 presents the results of the numerical experiments and discusses their importance for informing energy park design practices. An initial experiment quantifies the value of being able to update the energy park design, and change choice of energy storage technology, after uncertainty in storage technology performance has been reduced by R&D. However, as no information is available in the literature regarding how much R&D might reduce uncertainty, i.e. how closely the performance of a demonstrator system matches a large-scale version, a sensitivity analysis over the level of uncertainty reduction is performed. The VoO is then investigated in two further cases where two energy storage technologies are used in the park, and where risk aversion is incorporated into the design process. Finally conclusions are drawn in Section 6.4.

6.2 Experimental setup

6.2.1 Illustrative energy park system

The value of retaining optionality in energy storage technology choice during design is investigated for an illustrative energy park system. This system is modelled on a real-world energy park proposal. Due to its importance for decarbonizing transportation, heavy industry, and chemical manufacturing, there has been substantial commercial interest in the production of low-carbon hydrogen at scale. A proposal to build a 250 MW hydrogen electrolyser facility in the Port of Rotterdam to decarbonize industry and transportation in the area has received early-stage support from the Dutch government (BP, 2022; *Project - H2-Fifty 2022*). Co-locating renewable generation and energy storage with this industrial hydrogen plant in an energy park would reduce the total cost of operating the system, and its impact on the electricity grid.

6.2.2 System design under uncertainty

A key design challenge is determining the sizing of generation and storage, and the energy storage technology, that maximize the profitability of the energy park. It is assumed that offshore wind and solar generation (via direct wire connection) can be installed to support the site, and that energy can be traded with the grid. Fig. 6.1 illustrates the energy flows within the system model. Limits are imposed on grid connection capacity to reflect network restrictions, and on the maximum solar generation capacity to reflect land scarcity. Additionally, a capital budget constraint is imposed. Optimization-based design is used to select the energy storage technology and generation & storage sizings that minimize the expected cost of providing the industrial load. A carbon price is included to reflect the incentives for low-carbon hydrogen production. Uncertainties in the industrial load, patterns of wind and solar generation, and cost and technical performance of the different storage technologies are accounted for during design.

A two-stage decision model of the energy park design process is used. In an initial design stage, an energy storage technology is selected, and approximate sizings of the wind, solar, and storage capacities are determined (for consenting, planning permission, and procurement). At the time of initial design, the expected performance of each storage technology during operation is uncertain. As project planning and the pre-construction phases are completed, it is assumed that procurement and R&D is undertaken for the chosen storage technology, leading to the creation of a small-scale demonstrator. This demonstrator storage system provides the energy park designer with measurements of the storage technology performance, which reduce uncertainty in the actual performance of a

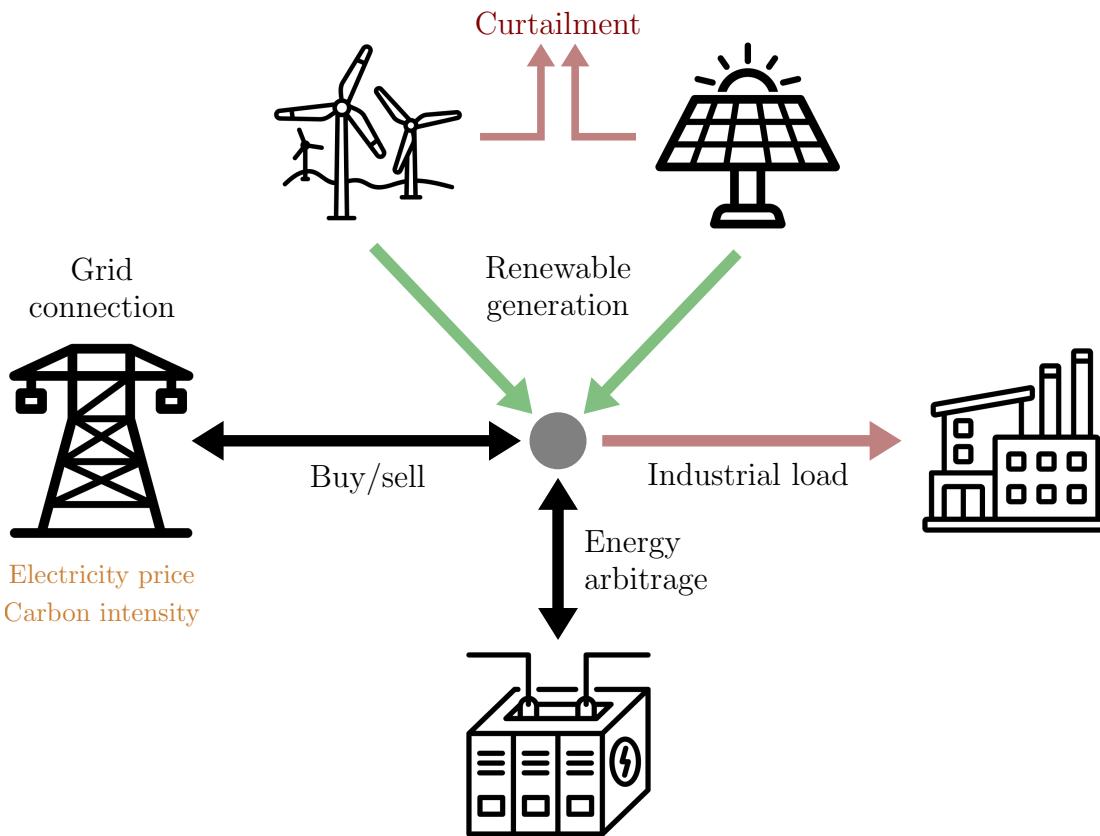


Fig. 6.1 Schematic of energy flows within energy park model

large-scale system. However, performing this procurement and R&D to reduce storage uncertainty is costly. After the demonstrator project is complete, a final design stage selects the exact wind, solar, and storage capacities for the energy park system. Finally, the system is constructed and operated, and the overall cost of providing the industrial load is observed. This decision model is illustrated in Fig. 6.2.

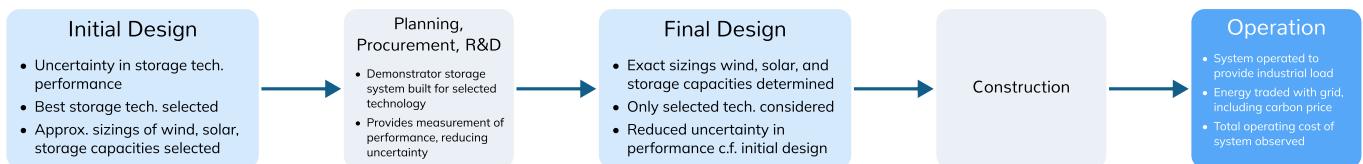


Fig. 6.2 Two-stage decision model of energy park design

This decision model raises a question,

Even though only one storage technology is ultimately used in the energy park, is it worthwhile performing procurement and R&D for all available technologies, so that at the final design stage, a possibly better choice of storage technology can be made once uncertainty has been reduced?

i.e.

Is retaining optionality in storage technology choice until the final design stage worth the cost?

6.2.3 Probabilistic models of renewable generation & industrial load

Historic data is used to construct probabilistic models of uncertain wind and solar generation. In both cases, generation potential per installed capacity data was gathered for 2010-2019. Uncertainty in renewable generation is modelled by randomly sampling a year from that range and using the corresponding generation data. For wind, data was collected from the [renewables.ninja](#) model (Staffell and Pfenninger, 2016a) at the location of a wind farm development zone near Rotterdam (IJmuiden Ver), using a power model from a typical offshore wind turbine. For solar, data was collected from the EU's [PVGIS](#) model (Commission, 2017), using default PV module settings. Fig. 6.3 plots this data from periods in both summer and winter for three example years.

Uncertainty in the demand for hydrogen in the local market during operation leads to uncertainty in the final sizing of the electrolyser. The conversion efficiency of the electrolyser units also has some uncertainty. So, at the time the supporting generation and storage infrastructure is designed, the electrical load that will be required by the hydrogen plant is uncertain. The industrial load is modelled as being constant during operation, as electrolyzers are typically run at high capacity factors to provide a competitive hydrogen price (Furfari and Clerici, 2021), but with an unknown/uncertain level. A truncated Gaussian distribution is used for the load, with mean set to the proposed electrolyser capacity of 250 MW, and standard deviation σ taken to be 25 MW (10%), truncated at $\pm 2\sigma$.

The probabilistic models of the renewable generation and industrial load time series are summarized in Table 6.1. Grid electricity price and carbon intensity data for the Netherlands from 2023 is used as a base case⁶⁰. Uncertainties in price and carbon intensity are not modelled, as due to global energy market events, few years of representative data are available. However 2024 data is used to perform a sensitivity analysis to determine if changes in grid price and carbon impact the conclusions derived.

⁶⁰ Further information on data collection and processing is available [online](#).

Parameter	Distribution	Parameters	Info
Wind	Discrete uniform	{2010, ..., 2019}	Years of historic generation potential data sampled
Solar			
Load	Truncated Gaussian	$\mu = 250 \text{ MW}$ $\sigma = 25 \text{ MW}$ cut-off = $\pm 2\sigma$	Truncation used to limit unrealistic extreme load values

Table 6.1 Probabilistic models of renewable generation and industrial load

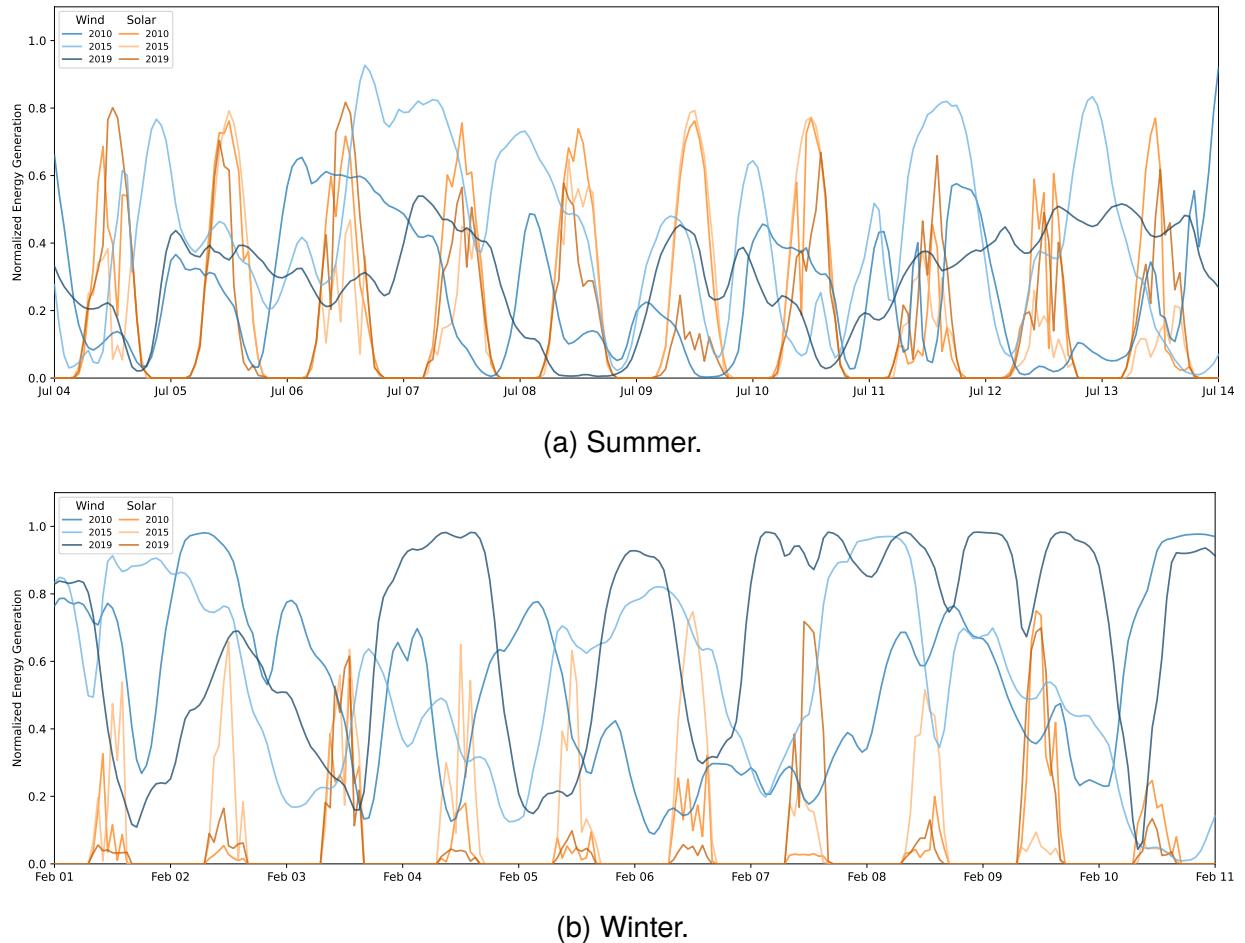


Fig. 6.3 Examples of wind and solar normalized generation profiles for Rotterdam.

The complete dataset of wind and solar generation profiles used in the case study can be viewed interactively [here](#).

6.2.4 Probabilistic models of storage technology performance & measurement

At the time the energy park is designed, the actual performance of the available storage technologies during operation (with regards cost and technical aspects such as efficiency) is unknown/uncertain. However, the energy system designer can develop a probabilistic understanding of how each storage technology might perform from existing storage systems (if any exist) and estimates from the scientific literature.

This study considers four candidate energy storage technologies for use in the energy park:

- Lithium ion batteries ([Li-ion](#))
- Sodium–Sulphur high-temperature batteries ([NaS](#))
- Vanadium Redox Flow battery ([VRFB](#))
- Compressed air energy storage ([CAES](#))

and considers uncertainty in the following economic and technical performance parameters of each storage technology:

- Cost per unit energy capacity (€/kWh)
- System lifetime (years)
- Round-trip efficiency (%)

It is assumed that the depth-of-discharge and discharge ratio⁶¹ of each technology are known, with values taken from ([IRENA, 2017](#)) and ([Kebede et al., 2022](#)) respectively.

The level of uncertainty varies both between different technologies, due to varying levels of technology maturity and research, and between performance parameters, due to the underlying factors causing uncertainty, such as raw material prices.

The uncertainty in each parameter for each technology at the time of initial design before any R&D has been performed, is modelled using a Gaussian distribution truncated at two standard deviations. This is referred to as the prior probabilistic model,

$$\theta \sim TN(\mu, \sigma, 2\sigma) \quad (6.1)$$

All uncertain parameters are taken to be independent, with mean (μ) and standard deviation (σ) values derived separately for each parameter from literature estimates.

⁶¹ The ratio of power capacity to energy capacity, also frequently referred to using the ‘storage duration’ (how long the storage unit can discharge for at full power output) which is the reciprocal.

When R&D is performed and the demonstrator system is built and tested, this provides the energy park designer with a measurement of the performance of that storage technology. This measurement can be used to update their understanding of how a large-scale storage system for that technology will perform. The information provided by this measurement is combined with the prior probabilistic model to produce a posterior distribution describing the remaining (reduced) uncertainty. The probability of obtaining a measurement, z , from the demonstrator system given the true performance of a large-scale storage system, θ , called the likelihood model, is taken to be,

$$z|\theta \sim \mathcal{N}(\theta, r\sigma) \quad (6.2)$$

The uncertainty reduction factor, r , models the confidence that the energy park designer has in the demonstrator, i.e. how closely they believe the performance of the demonstrator reflects that of a large-scale system. Determining a precise numerical value for this parameter is challenging, as there is very little data available regarding how the performance of demonstrator systems compares to large-scale energy storage plants, particularly in the case of newer technologies such as **CAES** and **VRFB** where few large-scale projects have been completed. An initial estimate of $r = 0.25$ is used for the experiments, representing a case where the designer has good confidence in the demonstrator, but retains a reasonable fraction of the initial uncertainty. To overcome this limitation, a sensitivity analysis over the value of r is performed. Note that because r is a multiplicative factor, the more mature storage technologies with lower uncertainty in their prior distributions also have correspondingly lower uncertainties in the posteriors.

Table 6.2 summarises the probabilistic models of the uncertain performance of large-scale storage systems, and the measurements obtained from the demonstrators, providing mean and standard deviation values for each uncertain parameter. Fig. 6.4 illustrates the prior distribution for the round-trip efficiency of **VRFB** storage, and the corresponding posterior distribution for an example measurement from a demonstrator system, showing the reduction in uncertainty provided by the demonstrator.

Probabilistic models					
Prior	$\theta \sim TN(\mu, \sigma, 2\sigma)$				
Likelihood	$z \theta \sim \mathcal{N}(\theta, r\sigma)$				
<i>Independent for each parameter & technology</i>					
Parameter	Units	Technology	Model parameters	Std. dev. σ	References
Cost	€/kWh	Li-ion	Mean μ	50	(IRENA, 2017)
		NaS	200	37.5	
		VRFB	175	75	
		CAES	250	15	
Lifetime	years	Li-ion	50	15	(Kebede et al., 2022)
		NaS	20	5	
		VRFB	25	5	
		CAES	20	2.5	
Efficiency	%	Li-ion	92	3.5	(Kebede et al., 2022)
		NaS	80	5	
		VRFB	75	5	
		CAES	60	2.5	

Table 6.2 Probabilistic models of uncertain storage technology performance parameters and their measurement.

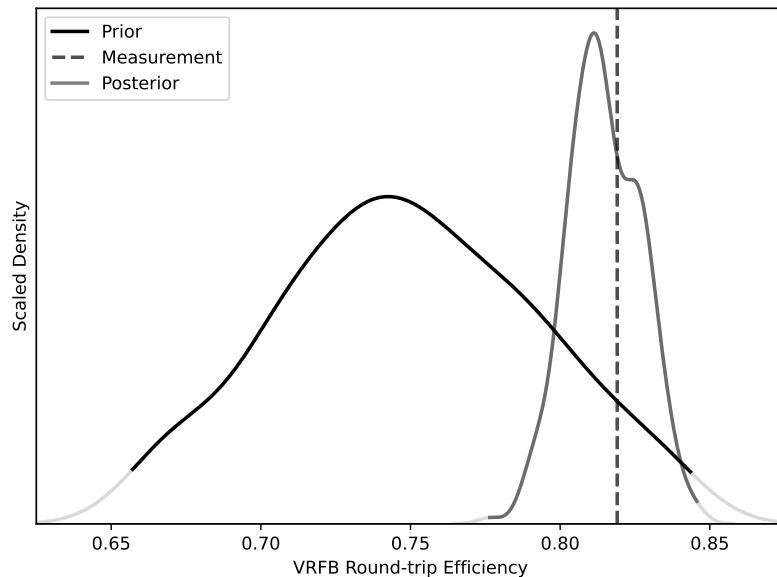


Fig. 6.4 Prior distribution of storage efficiency, and corresponding posterior for an example measurement value.

Note: distributions (kernel density estimates) are not quite truncated Gaussian due to small number of samples used (250); densities have been scaled to allow the distributions to be visually compared.

6.2.5 Stochastic Programming for system design

Stochastic Programming is used to optimize the capacities of wind and solar generation and energy storage in the energy park system. Specifically, linear scenario programming is used due to its computational efficiency, enabling the use of an hourly resolved model with multiple scenarios, and its resulting prevalent use in the literature for designing district- and national-scale energy systems (DeCarolis et al., 2017; Pickering and Choudhary, 2019; Yue et al., 2018).

A linearized model of energy flows within the park system (refer to Fig. 6.1) is used. The system design optimization task is formulated as a Linear Program⁶². Eq. 6.3 describes this formulation for general case where multiple storage technologies are used together in the energy park. Table 6.3 provides descriptions of all model parameters. To determine the best storage technology (or combination), the optimization is solved for each option independently, and the lowest cost setup is identified.

The design optimization aims to minimize the expected annualized cost of the energy park, the average total cost of running the energy park to provide the industrial load over all scenarios considered. The total cost is made up of the capital costs of the wind, solar, and energy storage, the cost (or profit) of energy bought from/sold to the grid, and the cost of carbon emissions associated with purchased grid electricity. The optimization is subject to energy conservation⁶³ (Eq. 6.3a), storage capacity (Eq. 6.3b & Eq. 6.3c), and grid capacity (Eq. 6.3h) constraints for operation within each scenario m . As well as a budget constraint on total capital expenditure (Eq. 6.3j). During operation, electricity can be bought from and sold to the electricity grid (Eq. 6.3f), and the renewable generation can be dynamically curtailed (Eq. 6.3g).

Scenarios are generated by sampling from the distributions of renewable generation, industrial load, and storage technology performance described in Sections 6.2.3 and 6.2.4. The sampled storage capacity cost and lifetime values are combined to produce annualized capacity cost samples for each technology, $p_{i,m}^s$. The number of scenarios that can be considered during optimization is limited by high computational cost. Therefore, scenario reduction (Pickering, 2019) is used⁶⁴ to improve the statistical representation of possible scenarios in the optimization. A large initial sample of scenarios is drawn, the optimization is performed considering each scenario separately, and a subset of scenarios which best represent the distribution of individually optimized costs is selected for use in the SP.

The parameter settings used across all experiments in this study are detailed in 6.A.

⁶² As the design optimization is solved approximately using a LP, estimates of the VoI and VoO are provided. The accuracy of such estimates is discussed in Chapter 5.

⁶³ $[\cdot]^+$ and $[\cdot]^-$ represent the positive and negative parts of the argument respectively.

⁶⁴ Fast-forward scenario reduction algorithm from (Heitsch and Römisich, 2003), implemented by (Gioia, 2023), used.

$$\min \quad p^w C^w + p^{pv} C^{pv} + \sum_m \rho_m \left(\sum_i p_{i,m}^s C_i^s + \sum_t \left(p^e[t] E_m^{\text{grid}}[t] + p^c c[t] [E_m^{\text{grid}}[t]]^+ \right) \right) \quad (6.3)$$

over $C^w, C^{pv}, C_i^s, \nabla_m[t], E_{i,m}[t], \text{SoC}_{i,m}[t+1] \quad \forall i, m, t$

subject to $\text{SoC}_{i,m}[t+1] = \text{SoC}_{i,m}[t] + \sqrt{\eta_{i,m}} [E_{i,m}[t]]^+ - 1/\sqrt{\eta_{i,m}} [E_{i,m}[t]]^- \quad (6.3a)$

$$-P_i^{\max} \Delta t \leq E_{i,m}[t] \leq P_i^{\max} \Delta t \quad (6.3b)$$

$$(1-\nu_i) C_i^s \leq \text{SoC}_{i,m}[t+1] \leq C_i^s \quad (6.3c)$$

$$P_i^{\max} = \delta_i C_i^s \quad (6.3d)$$

$$\text{SoC}_{i,m}[0] = \text{SoC}_0^0 C_i^s \quad (6.3e)$$

$$E_m^{\text{grid}}[t] = L_m[t] - (C^{pv} g_m^{pv}[t] + C^w g_m^w[t] - \nabla_m[t]) + \sum_i E_{i,m}[t] \quad (6.3f)$$

$$\nabla_m[t] \leq C^{pv} g_m^{pv}[t] + C^w g_m^w[t] \quad (6.3g)$$

$$E_m^{\text{grid}}[t] \leq C^{\text{grid}} \Delta t \quad (6.3h)$$

$$C^{pv} \leq C_{\max}^{pv} \quad (6.3i)$$

$$p^{pv} C^{pv} + p^w C^w + \sum_i p_{i,m}^s C_i^s \leq B^{\text{cap}} \quad (6.3j)$$

for all $i \in [0, I-1], m \in [0, M-1], t \in [0, T-1]$

Parameter	Units	Description
Decision variables		
C^w	kWp	Installed offshore wind generation capacity
C^{pv}	kWp	Installed solar PV generation capacity
C_i^s	kWh	Installed energy capacity of storage technology i
$\nabla_m[t]$	kWh	Generation curtailment at time t in scenario m
$E_{i,m}[t]$	kWh	Energy <i>intake</i> to storage technology i at time t in scenario m
$SoC_{i,m}[t]$	kWh	State-of-charge of storage technology i at time t in scenario m
Derived variables		
$E_m^{\text{grid}}[t]$	kWh	Net energy <i>drawn from</i> electricity grid at time t in scenario m
P_i^{\max}	kW	Power capacity of storage technology i
Sampled parameters		
ρ_m	—	Probability of scenario m
$L_m[t]$	kWh	Industrial electrical load at time t in scenario m
$g_m^w[t]$	kWh/kWp	Wind power generation potential at time t in scenario m
$g_m^{pv}[t]$	kWh/kWp	Solar PV power generation potential at time t in scenario m
$\eta_{i,m}$	—	Round-trip efficiency of storage technology i in scenario m
$p_{i,m}^s$	€/kWh/yr	Annualized capacity cost [†] of storage technology i in scenario m
Known parameters		
Δt	hrs	Time step of simulation data
ν_i	—	Depth-of-discharge of storage technology i
δ_i	kW/kWh	Discharge ratio of storage technology i (power capacity/energy capacity)
SoC^0	—	Initial state-of-charge of storage (fraction of capacity)
p^w	€/kWp/yr	Annualized capacity cost [†] of offshore wind
p^{pv}	€/kWp/yr	Annualized capacity cost [†] of solar PV
$p^e[t]$	€/kWh	Price of grid electricity at time t
p^c	€/kgCO ₂	Nominal carbon price
$c[t]$	kgCO ₂ /kWh	Carbon intensity of grid electricity at time t
C^{grid}	kW	Grid connection capacity
C_{\max}^{pv}	kWp	Solar capacity limit
B^{cap}	€/yr	Annualized capital budget constraint
Indices		
i	—	Storage technology
t	—	Time step in modelled operation
m	—	Scenario number

Table 6.3 Description of Stochastic Program variables & parameters.

[†] Capital cost per energy/power capacity per year of lifetime

Accounting for risk aversion

The optimization objective in Eq. 6.3 assumes that the energy park designer is risk-neutral and is only interested in minimizing the average operating cost of the system, as is standard in the energy system design literature. However, when investing billions of Euros in large energy infrastructure projects, energy firms are typically risk-averse in their decision making, and are concerned about the risk of high operating costs. Previous studies have investigated the impact of risk aversion on energy system design (Mu et al., 2022; Pickering, 2019; Tostado-Véliz et al., 2024), but none have considered how risk aversion affects the value of uncertainty reduction or optionality during decision making.

To account for the impact of risk aversion during energy park design, a risk measure is included in the objective function. The **Conditional Value-at-Risk (CVaR)** is used as it is a coherent risk measure compatible with Linear Programming (Rockafellar and Uryasev, 2000). For a linear scenario program where the cost and probability of each scenario are c_m and ρ_m respectively, an optimization minimizing the CVaR is given by,

$$\begin{aligned} \min \quad & \xi + \frac{1}{\alpha} \sum_m \rho_m \eta_m \\ \text{s.t.} \quad & \eta_m \geq 0 \text{ and } \eta_m \geq c_m - \xi \end{aligned} \tag{6.4}$$

When optimized, this objective is equal to the expected value of the costs in the α right-tail of scenarios, $\mathbb{E}_{m: c_m > \xi_\alpha} \{c_m\}$, i.e. the average of the highest $\alpha\%$ of costs across the scenarios.

As **Vol** and **VoO** are differences in average costs, to allow valid comparison between the risk-neutral and risk-averse cases, we develop a risk-averse objective which maintains the scale of the cost. This objective is equivalent to a weighted average of the scenario costs, where the costs in the $\alpha\%$ right-tail are weighted by a factor $n+1$ relative to others. So, risk aversion is represented by the designer ‘caring’ about high operating costs more than others, meaning the objective is still a physical cost, and so can be validly compared across cases. This risk-averse objective is,

$$\frac{1}{1+n\alpha} \left(\underbrace{\mathbb{E}\{c_m\}}_{\text{original obj.}} + n\alpha \underbrace{\left(\xi + \frac{1}{\alpha} \sum_m \rho_m \eta_m \right)}_{\text{CVaR}} \right) \tag{6.5}$$

where c_m is the objective value in each scenario as specified in Eq. 6.3, and both the original constraints for Eq. 6.3 and the constraints from Eq. 6.4 are imposed. When optimized this objective becomes $(\mathbb{E}\{c_m\} + n\alpha \mathbb{E}_{m: c_m > \xi_\alpha} \{c_m\}) / (1 + n\alpha)$.

6.3 Results & Discussion

6.3.1 Prior design

Initial designs for the energy park using each storage technology are produced by applying the Stochastic Program (SP) model (described in Section 6.2.5), using the prior distributions of renewable generation, industrial load, and storage technology performance (defined in Sections 6.2.3 and 6.2.4). Table 6.4 compares the performance of these designs. Sodium–Sulphur high-temperature batteries (NaS) are found to provide the lowest total system cost, so this design is selected, and NaS is chosen for procurement and R&D. The design sizes 441 MW wind and 500 MW solar generation capacity, with 1.51 GWh of NaS storage to support the system. The average total cost of this system is €124.1m/yr, with €10.7m/yr expected to be spent on batteries. If no energy storage were used in the energy park, 486 MW of wind and 500 MW of solar generation would be installed, leading to an average cost of €183.7m/yr. Therefore, installing bulk energy storage for arbitrage reduces the average cost of the energy park system by €59.6m/yr (32%), and reduces the average operational carbon emissions by 20.9 ktCO₂/yr (20%).

Storage technology	Total cost (€m/yr)	Carbon emissions (ktCO ₂ /yr)	Storage capacity (GWh)	Cost of storage (€m/yr)
None	183.7	104.3	–	–
CAES	148.1	84.2	3.59	7.2
Li-ion	150.1	87.5	0.93	9.2
NaS	124.1	83.4	1.51	10.7
VRFB	176.6	93.9	0.40	5.2

Table 6.4 Performance of initial energy park designs using different storage technologies.

Cost and carbon values are averages over prior distributions of load, generation, and storage performance.

6.3.2 Design updating without optionality & the Value of Information

To investigate the final design of the energy park given the uncertainty in the performance of NaS storage, 250 possible measurement values obtained from the demonstrator system built during R&D were sampled for each performance parameter using the probabilistic model of storage performance (see Table 6.2). For each sampled measurement, the SP model is applied using the corresponding posterior distribution of large-scale system performance to produce a final design. Fig. 6.5 plots the wind generation and NaS storage

capacities of the final designs for each sampled measurement, and compares them to the initial design. All designs install the maximum 500 MW of solar.

The average total cost of the final energy park design over the samples is €102.0m/yr, whereas the cost of the initial design was €124.1m/yr. Therefore, the [Value of Information \(Vol\)](#) associated with designing the energy park after R&D has been performed to reduce uncertainty in the performance of [NaS](#) storage is €22.1m/yr (18% of initial design cost). This cost reduction comes from the system designer's improved ability to trade off wind generation and storage capacity depending on the relative cost of energy arbitrage, and better make use of the capital budget. In some cases where the cost of [NaS](#) storage is low and its efficiency is high, the storage capacity in the final design is more than double that initially chosen. Assuming a 20 year project lifetime, having the option to update the capacities of wind and storage procured from their suppliers after R&D is worth €442m to the energy park developer. As a result, they should be willing to pay more (per unit capacity) to a supplier offering a flexible contract.

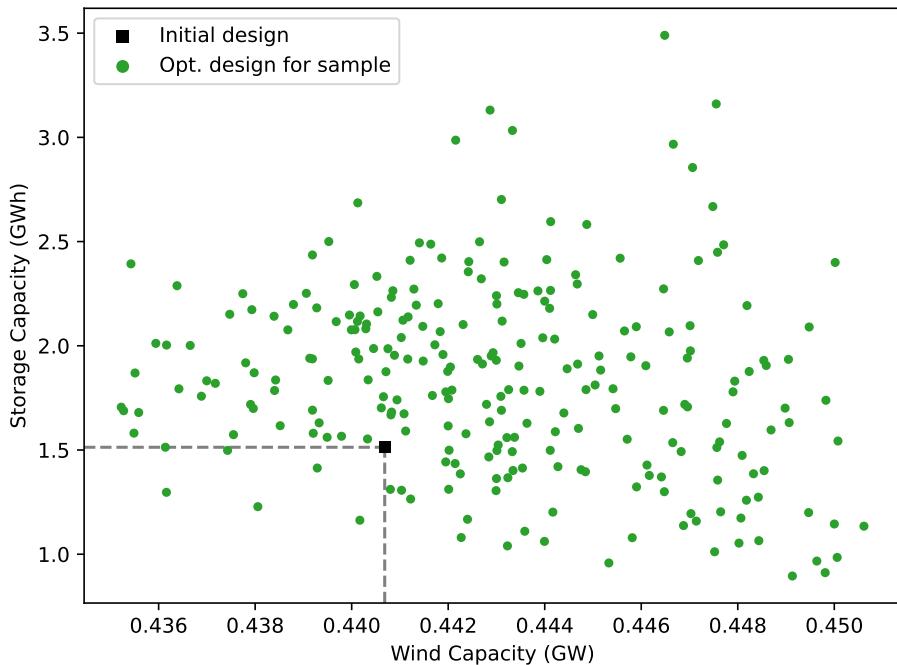


Fig. 6.5 Final designs of energy park with [NaS](#) storage for each sampled measurement.

6.3.3 Designing with optionality & the Value of Optionality

In the initial design, NaS was found to provide the lowest total cost *on average*, and R&D was performed only for this technology. However, it may be the case that NaS does not end up being the best technology, and after storage performance uncertainty is reduced another technology provides a lower average cost. Fig. 6.6 plots the distribution of the expected costs of final designs for each storage technology, obtained by repeating the process of sampling measurement values and optimizing sizings (as described in the previous section). The significant overlap in the distributions indicates there are likely several cases where other storage technologies provide a lower average cost than NaS once performance uncertainty has been reduced.

The average cost of the energy park could therefore be reduced by performing R&D with all storage technologies, allowing the best technology to be selected during final design when uncertainty has been reduced. However, performing R&D, developing demonstrator systems, and purchasing supplier contracts giving the right-to-purchase for all energy storage technologies comes at a significant cost.

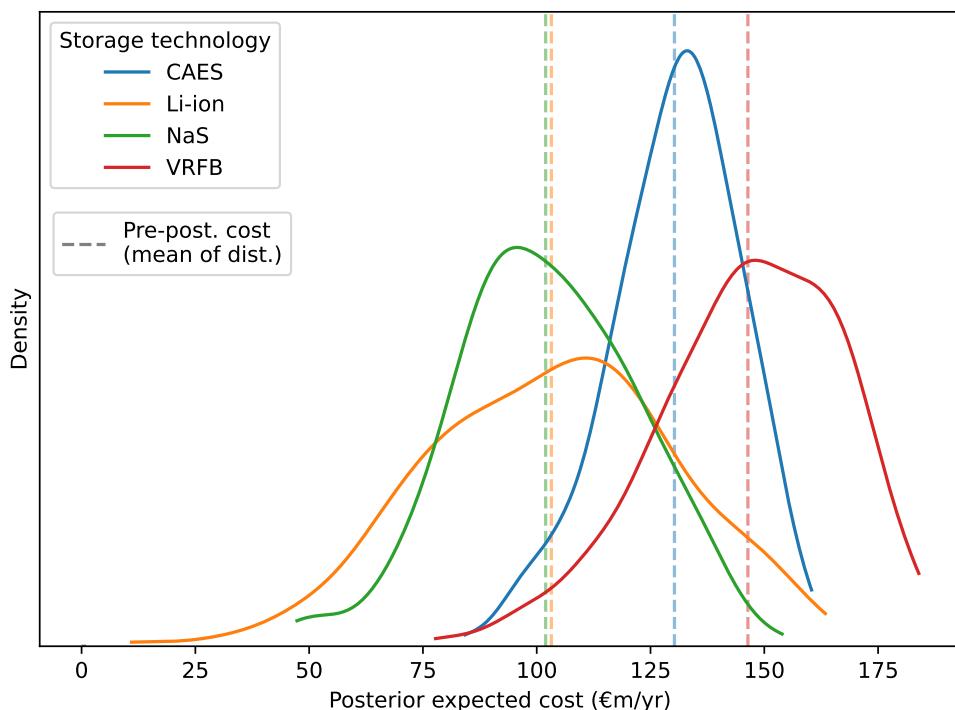


Fig. 6.6 Distribution of final design expected costs over sampled measurements for each storage technology.

Dashed line indicates total cost of final designs averaged over sampled msrmnts., which is Pre-Posterior cost (Eq. 2.10).

The process of sampling possible demonstrator measurement values is repeated (for all technologies), and final energy park designs are determined for each sample, this time allowing the best technology to be selected. Fig. 6.7 plots the resulting final designs for each sample. The marker colour indicates which storage technology is selected in each case. In 132 of 250 samples (53%) a different technology than **NaS** is selected for the final design. In 123 of those cases **Li-ion** is the best technology choice after performance uncertainty has been reduced. Fig. 6.8 illustrates for each sample, how much the average cost of the energy park is reduced by being able to select the best storage technology, compared to the case where only **NaS** is available.

The average cost of the energy park designed *with* storage technology optionality is €88.7m/yr, compared to €102.0m/yr if only **NaS** is available for final design. Therefore, the value of retaining storage optionality for the final design of the energy park (**VoO**, discussed in Section 2.5.2) is €13.3m/yr (13% of cost without optionality). With a 20 year system lifetime, this optionality in storage technology selection is worth €266m to the energy park developer. Long-duration energy storage demonstrator projects funded by the UK government have received around €1m for proof-of-concept systems (DESNZ, 2023c; Inc, 2023) and €10m for mid-scale demonstrators (DESNZ, 2023d,e). So, the cost of developing and contracting **Li-ion**, **VRFB**, and **CAES** storage is conservatively estimated at €20m each. Therefore, hedging bets on storage technology performance and maintaining optionality in technology choice through the energy park design process provides a net reduction in average cost of €206m (10%) to the energy park developer.

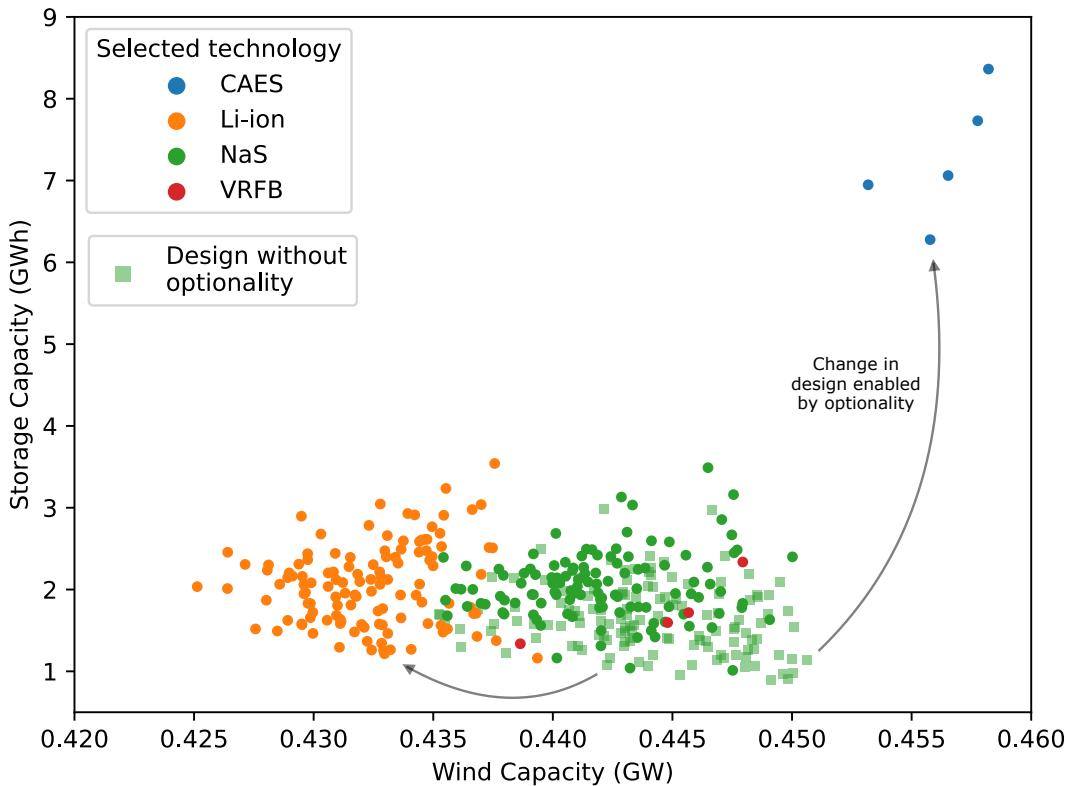


Fig. 6.7 Final designs of energy park with storage optionality for each sampled measurement.

Marker colour indicates the selected storage technology. Squares indicate final designs using NaS storage for samples where NaS does not provide the lowest average cost.

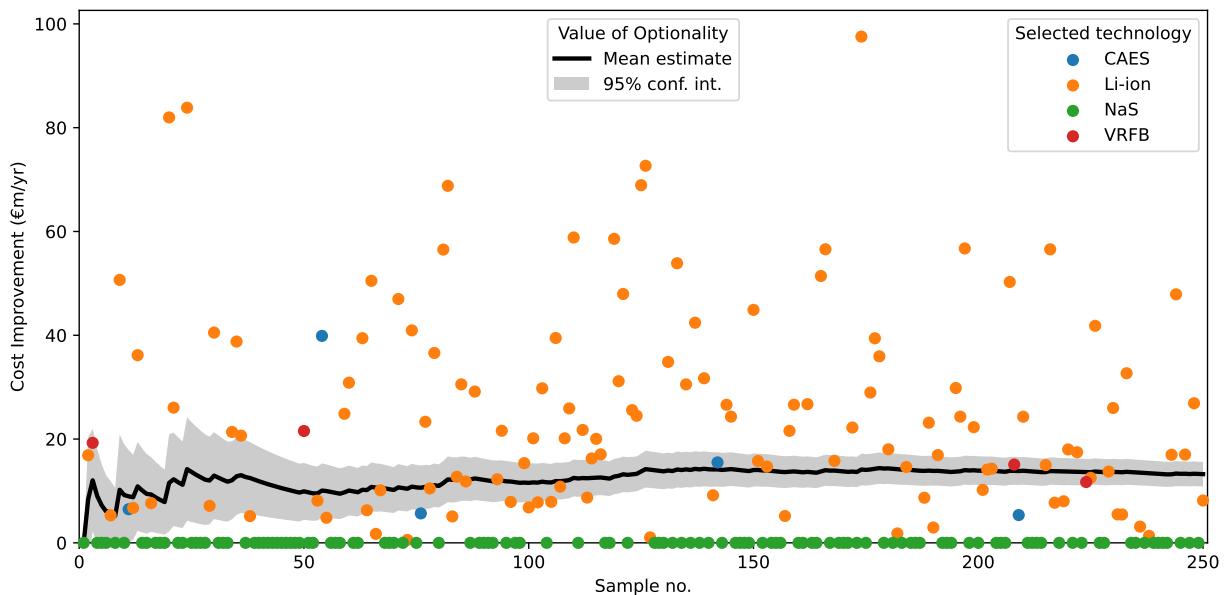


Fig. 6.8 Reduction in average cost of final design provided by storage technology optionality, and selected technology, for each sampled measurement.

6.3.4 Sensitivity analysis

A key assumption made in the probabilistic model of uncertain storage performance in Section 6.2.4 was the level of uncertainty reduction provided by measuring the demonstrator system built during R&D. This is encapsulated by the uncertainty reduction factor parameter r in Eq. 6.2, which quantifies how closely the designer believes the performance of the demonstrator matches that of a large-scale system, and so how much uncertainty remains after measurement. Finding data from real-world storage systems to estimate this parameter is extremely challenging, as actual performance and cost data are highly commercially sensitive, and few large-scale systems currently exist.

Sensitivity analysis is performed to determine whether the conclusions provided by the **Value of Optionality** analysis remain valid. The process from Sections 6.3.2 and 6.3.3 is repeated for values of r over the feasible range. Table 6.5 shows **VoI** and **VoO** values for each case. For both **VoI** and **VoO**, the greater the uncertainty reduction provided by measurement (lower r), the greater the benefit to design. This is because lower uncertainty allows final designs with less hedging to be used, which achieve lower cost. While there will be some threshold of r above which the **VoO** is less than the cost of R&D, even with the lowest level of uncertainty reduction tested, $r = 0.5$, retaining storage technology selection optionality is still economically worthwhile. So the conclusions of the base case analysis are robust to the level of uncertainty reduction. The variation of **VoI** with r is larger than that of **VoO**, as **VoO** compares differences in designs both made with reduced uncertainty, so both costs in the difference benefit (reduce) from lower uncertainty.

During operation the energy park is able to trade with the grid to reduce the cost of the energy it supplies to the hydrogen plant, but must pay a cost for any associated carbon emissions. However, due to lack of data availability, uncertainty in grid electricity price

Uncertainty reduction factor, r	Value of Information (€m/yr)	Value of Optionality (€m/yr)
0.1	26.4	15.1
0.2	23.5	13.9
0.25	21.1	13.3
0.3	20.6	12.9
0.4	18.5	11.6
0.5	16.6	9.8

Table 6.5 Sensitivity of **Value of Information** and **Value of Optionality** to level of uncertainty reduction.

Case	Prior operating cost (€m/yr)	Value of Information (€m/yr)	Value of Optionality (€m/yr)
2023 prices & carbon (base case)	124.1	22.1	13.3
2023 prices +10%	124.9	39.2	16.1
2023 prices -10%	151.5	32.1	12.1
2024 prices & carbon	80.9	38.2	17.1

Table 6.6 Sensitivity of Value of Information and Value of Optionality to electricity prices and carbon emissions

and carbon emissions were not included in the uncertainty model. So, a similar sensitivity analysis is performed to check whether the pricing and carbon data assumed in the base case are representative, or whether changes in this data have a significant effect on the resulting VoI and VoO values.

Three alternative cases are tested: the 2023 pricing data is increased and decreased by 10% respectively, and pricing and carbon data from 2024 is used. Table 6.6 presents the VoI and VoO values for each of these cases. While the VoI is found to vary significantly, by up to 75%, in all of the alternative cases it is greater than the base case. This suggests that design flexibility could be even more valuable than indicated by the initial results. As before, the changes in VoO values are smaller than for the VoI. And again, in all cases the VoO is greater than the cost of developing the additional storage technologies, so retaining optionality is still found to be economically worthwhile. Therefore, the conclusions from Sections 6.3.2 and 6.3.3 are not affected by the electricity price and carbon emissions data used.

Whether or not retaining optionality in storage technology selection is worthwhile is determined by the comparison between the VoO and the cost of retaining optionality. The cost of optionality is therefore a critical driver of this conclusion. Were the €20m cost of building a demonstrator system used in the initial analysis (see Section 6.3.3) to double, this would correspond to an annualized cost of €6m/yr for developing all three additional storage technologies. However, even this doubling of a pessimistic cost estimate is still significantly smaller than any of the VoO values found across the sensitivity analyses, the smallest of which is €9.8m/yr from the extreme case of low uncertainty reduction, $r = 0.5$. So, the assumptions made about the cost of optionality do not affect the conclusions drawn from the analysis, as the VoO is sufficiently large to provide strong confidence that storage optionality is worthwhile for the design of the studied energy park.

6.3.5 Use of multiple storage technologies

So far it has been assumed that only a single energy storage technology is installed in the energy park to perform arbitrage. The different characteristics of the energy storage technologies (such as efficiency, cost per energy capacity, and discharge ratio) make them suited for arbitrage over different time scales (IRENA, 2017). As trends in renewable generation and grid electricity price also occur over a range of time scales, the profit from energy arbitrage could be increased by installing multiple energy storage technologies, and sizing them for the volume of the arbitrage they are best suited to.

The investigation is repeated for the case where two storage technologies are used in the energy park. This means two technologies are selected and developed following the initial design. Fig. 6.9 plots an example of the operation of the energy park with two storage technologies, for the initial system design with Li-ion and CAES. The state-of-charge traces for Li-ion (orange) and CAES (dark blue) demonstrate the different arbitrage behaviours of the two technologies. As the Linear Program simulating the energy park (Eq. 6.3) simultaneously optimizes over all operational variables, it is able to dynamically dispatch the two storage technologies, scheduling them to best suit their characteristics and maximize the combined performance. In this case, best performance is achieved by operating the Li-ion at short time scales with small traded energy volumes, while the CAES is operated over longer time scales with greater energy volumes.

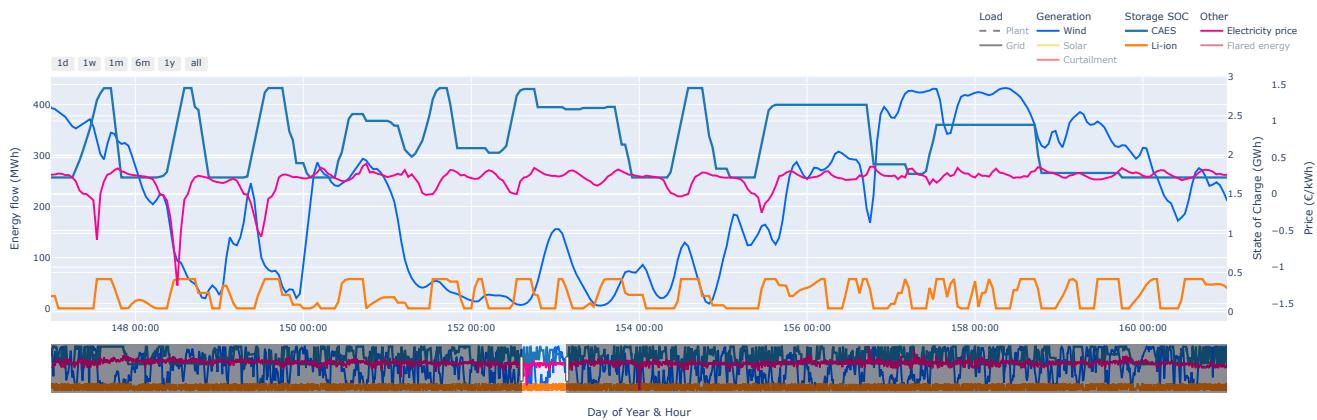


Fig. 6.9 Operation of energy park with Li-ion and CAES storage, demonstrating different energy arbitrage time scales.

Interactive version of plot available online [here](#).

For the energy park with two technologies, the average cost of the final designs is €87.5m/yr for the case without optionality, and €85.7m/yr for the case with optionality. Table 6.7 compares these costs to the results for the system with only one storage technology from Sections 6.3.2 and 6.3.3. While using two storage technologies provides a substantial cost reduction for the case without optionality in the final design, €14.5m/yr (14%), with optionality this cost reduction falls to €3.0m/yr (3%). However, this cost saving is still significant (€60m over a 20yr system lifetime), and shows that multiple energy storage technologies should be used for arbitrage in energy parks.

Design case	Total cost (€m/yr)	Vol/VoO (€m/yr)	Carbon emissions (ktCO ₂ /yr)	Cost of storage (€m/yr)
No storage	183.7	–	104.3	–
1 tech.	Initial design	124.1	–	10.7
	Design with flexibility*	102.0	22.1	13.0
	Design with optionality	88.7	13.3	14.4
2 techs.	Initial design	114.0	–	11.2
	Design with flexibility*	87.5	26.5	15.1
	Design with optionality	85.7	1.8	15.0

Table 6.7 Comparison of energy park performance for different design cases.

*All values are averages over the relevant distributions. *Updated (final) designs without optionality.*

The VoO is now €1.8m/yr (2.1%), or €36m over the 20 year lifetime. This is less than the estimated €40m cost of developing the remaining two storage technologies not selected in the initial design. So, when two technologies are used, it is not worthwhile hedging bets against uncertain storage performance by retaining optionality. This means selecting the best two storage technologies is robust to the performance uncertainty. The Vol however remains high, €26.4m/yr, meaning that contract flexibility and updating the final system design are still very valuable.

6.3.6 Risk-averse design

The objective function used to design the energy park in previous sections, Eq. 6.3, assumes that the designer is risk-neutral and is only interested in the average system cost. However, for capital-intensive energy infrastructure projects at this scale, energy firms are usually risk-averse and concerned by the possibility of poor financial returns. The effect of risk aversion on the energy park design, and the benefits of uncertainty reduction (**Vol**) and optionality (**VoO**) for improving design, are studied by repeating the investigation using the risk-averse objective defined in Eq. 6.5, which increases the weighting on the worst $\alpha\%$ of scenarios n -fold in the cost average.

For the case where $\alpha = 10\%$ and $n = 1$, adding risk aversion increases the expected cost of the initial design from €124.1m/yr in the risk-neutral base case to €136.6m/yr, with **NaS** again being selected as the best storage technology. This cost increase occurs as the system designer is now more pessimistic, and puts more weighting on the highest 10% of costs across the scenarios. However, whilst the expected cost increases by 10%, the system sizing remains unchanged. The same pattern is found for the final designs, with costs increasing by 12.5% on average but the designs changing negligibly.

Table 6.8 presents the **Vol** and **VoO** values calculated for various combinations of α and n . Increasing risk aversion causes a small decrease in **Vol**, whereas the impact on **VoO** is negligible. This indicates that the cost variability for the final designs, and so the risk penalty, is the same across the storage technologies. So this variability must be driven by the uncertainties in renewable generation and industrial load (see Table 6.1) rather than the storage uncertainties. Therefore, while reducing storage uncertainty is important for system design, it does not have a significant impact on risk.

Confidence level, α (%)	Tail-weighting, n	Vol (€m/yr)	VoO (€m/yr)
10	1	21.8	13.3
	2	21.5	13.4
	5	20.8	13.5
25	1	22.1	13.4
	2	22.0	13.5
Base case (no risk aversion)		21.1	13.3

Table 6.8 **Value of Information** and **Value of Optionality** for risk-averse design cases

6.4 Conclusions

This chapter investigated the impact of uncertainty in the cost, lifetime, and efficiency of energy storage technologies on energy park design, and the benefit of retaining optionality in storage technology choice through the design process. It quantified the improvement in design (reduction in average total cost) achieved when multiple technologies are developed to retain optionality in technology choice, and allow a better storage selection to be made after uncertainty in performance has been reduced. An illustrative energy park system containing wind and solar generation and bulk energy storage to support a hydrogen electrolyser, modelled off a real-world proposal in the Port of Rotterdam, was used for experiments. Four candidate bulk energy storage technologies were considered: Lithium-ion batteries ([Li-ion](#)), Sodium–Sulphur high-temperature batteries ([NaS](#)), Vanadium redox flow batteries ([VRFB](#)), and compressed air energy storage ([CAES](#)).

For the base case where only a single storage technology was selected for development, updating the system design (wind, solar, and storage sizings) after R&D, when an improved estimate of storage performance is available (i.e. with reduced uncertainty), was found to reduce total costs by 18% on average ([VoI](#) of €442m). This demonstrates that flexibility in procurement contracts for generation and storage is highly valuable to energy park developers.

If all four energy storage technologies were developed, so that when the system design was updated any technology could be selected for the final design, the average total cost was found to reduce by 13% compared to the case without optionality ([VoO](#) of €266m). As the cost of performing R&D for three additional storage technologies was estimated at €60m, retaining optionality in storage technology selection through the design process is economically worthwhile, and reduces the net cost of the energy park project by €206m on average.

Installing two storage technologies in the energy park to perform energy arbitrage at different time scales reduced the total system cost by 14% on average (€290m) compared to the base case without optionality. And in this, case developing the two extra storage technologies to provide optionality was determined to not be worthwhile, reducing average costs by €36m ([VoO](#)) compared to the €40m additional cost of R&D. The [VoI](#) however increased to €528m, so updating the system design is more valuable when trade-offs can be made between *two* storage technologies and generation.

The decision-making recommendations from the [VoI](#) and [VoO](#) results were shown to be robust to the level of uncertainty reduction in energy storage performance provided by R&D used in the statistical model, as well as the grid electricity price and carbon emissions data used, and the cost of developing demonstrator storage systems assumed.

Additionally, including risk aversion in the design objective was found to have little impact on the optimized system designs or [VoI](#) and [VoO](#) results, as the risk (variability in system cost) was driven mostly by uncertainty in renewable generation and industrial load.

These results provide two important insights for energy park developers. Firstly, updating system designs after improved estimates of storage technology performance have been gathered (i.e. uncertainty reduced) provides substantial savings in system cost. Hence, flexible procurement contracts should be negotiated to allow design adjustments to be made as better information becomes available. Secondly, two energy storage technologies should be used to support the energy park. Doing so both significantly reduces the total cost of the system, and means that a robust choice of storage technologies can be made up-front, as hedging against the chance of a different technology providing better performance is not worthwhile if two technologies are developed and installed.

However, as the [VoI](#) framework does not provide generalisation guarantees, further work is required to determine whether these insights remain valid for other energy park systems, for example those with different industrial loads, different generation mixes, or in geographic locations with different weather patterns and so renewable generation uncertainties. The framework should also be used to study the importance of reducing uncertainty and providing optionality in other components of an energy park. For instance renewable generation technologies, where uncertainty in their energy production is driven by technical parameters and local site characteristics, which could be reduced by equipment testing and site surveys, that may then influence the site selection decision.

The results of this chapter demonstrate the importance of performing this type of uncertainty analysis using the [VoI](#) framework when designing energy parks. As they show that the total cost of an energy park system can be substantially reduced (in this case by 30%) by enabling the system design to be updated with optionality in storage technology choice after uncertainty in the performance of those storage technologies has been reduced. More broadly, these results also demonstrate the importance of considering the options available in engineering design, and the effect that uncertainties which can be reduced over the course of the design process have on initial design decisions. The [VoO](#) allows us to understand when we can make ‘good enough’ early-stage design decisions, and when it’s worth delaying design commitments and keeping optionality until better information is available.

Appendix 6.A Parameter settings used in experiments

Parameter	Units	Value	Note/Refs
Simulation duration, T	Hours	8760	
Simulation time step, Δt	Hours	1	
Initial state-of-charge, SoC ⁰	–	0.75	
Offshore wind CAPEX	€/kWp	5000	(Stehly et al., 2023), used to derive annualized capacity cost, p^w
Offshore wind OPEX	€/kWp/yr	100	"
Offshore wind lifetime	Years	20	"
Solar PV CAPEX	€/kWp	5000	(Ramasamy et al., 2023), used to derive annualized capacity cost, p^{pv}
Solar PV OPEX	€/kWp/yr	100	"
Solar PV lifetime	Years	20	"
Carbon cost, p^c	€/kgCO ₂	1	Chosen to be significantly larger than current carbon trading prices, c. €0.1/kgCO ₂ (DESNZ, 2022), (Bloomberg, 2024), to reflect incentive to produce green hydrogen.
Grid capacity, C^{grid}	MW	500	
Solar capacity limit, C_{\max}^{pv}	MW	500	
Capital budget	€m/yr	200	Corresponds to €4bn over a 20 year lifetime (Auchincloss, 2024)
<i>Storage parameters</i>			
Depth-of-discharge, ν_i	%	90	Li-ion, taken from (IRENA, 2017)
		100	NaS, "
		100	VRFB, "
		40	CAES, "
Discharge ratio, δ_i	kWp/kWh	2	Li-ion, taken from (Kebede et al., 2022)
		1	NaS, "
		0.5	VRFB, "
		0.1	CAES, "

Table 6.9 Parameter values for energy system model used in experiments.

Parameter	Value
No. samples from prior distribution	250
No. samples from each posterior distribution	250
MCMC sampling burn-in period	250
MCMC sampling thinning factor	10
No. of reduced scenarios used in Stochastic Program	25

Table 6.10 Settings for sampling from probabilistic models used in experiments.

7

Conclusions

We can only make good decisions with good data.
So we need to understand how the data we have affects our decision making, and what good data looks like.

This thesis explores the role of data for supporting decision making in energy systems. Its main aim was to quantify the improvement in decision making that data collection provides for the design and operation energy systems. This was then used to determine whether collecting data to inform these decisions is worthwhile, and if so how much data should be collected, as well as how the decision making process should be adapted to best use this data and the uncertainty reduction it provides. The computational experiments in this thesis investigate these questions of data collection requirements and the importance of uncertainty reduction for decision making across a broad range of decision making tasks, spanning design, operation, and management, and covering energy system scales from individual buildings to grid-scale energy parks.

A literature review demonstrated that while modelling and accounting for uncertainties during decision making is common practice in the energy systems field, little attention has been paid to how reducing these uncertainties can improve decision making. Similarly, while the importance of collecting data from energy systems is fully accepted, and its practical issues widely acknowledged, few studies have looked at the data collection requirements of energy systems, and many costs associated with data collection are frequently ignored. No existing study has quantified the benefit that uncertainty reduction from practical data collection would provide to decision making in energy systems, and determined whether this benefit is greater than the cost of collecting the data, i.e. whether the data collection is worthwhile.

Chapter 2 demonstrated how [Value of Information](#) analysis ([Vol](#)) can be used to address this research gap, as it provides a numerical framework for quantifying the improvement in decision making provided by uncertainty reduction. The theory and operation of the [Vol](#) calculations were explained, and alongside this a simple example problem involving the sizing of solar panels and battery storage in a commercial building was developed to illustrate how [Vol](#) can be applied to decision making in energy systems.

The capabilities of [Vol](#) to inform decision making about data collection were further demonstrated in Chapter 3. It explored three example decision problems in building energy systems: air source heat pump maintenance scheduling, ventilation scheduling for indoor air quality, and ground-source heat pump system design. In each example a different aspect of how [Vol](#) can be used to support decision making was illustrated. Across the three examples, there was a large variation in the benefit that each of the data collection options considered provided to decision making. Knowing the [ASHP](#) performance and degradation reduced its operation cost by just 0.06% on average through improved maintenance scheduling. Whereas knowing the office occupancy and improving ventilation scheduling reduced operating costs by 15% on average. While knowing the ground thermal conductivity to help select the best borehole length for a [GSHP](#) system reduced the lifetime cost by around 4% on average, and in this case a ground thermal test with a moderate precision was found to be most economical. This variation in [Vol](#) results highlights the importance of performing [Vol](#) analysis, as the value of data collection for supporting decision making is often not obvious even to those with expert knowledge of the energy system.

However, it's not always possible to use [Vol](#) when investigating the benefits of data collection and data requirements. Chapter 4 studied the impact of data on the accuracy of machine learning based forecasting models for building operational conditions, and the resulting operational performance when they are used in a [Model Predictive Control](#) scheme to control distributed generation and storage in a district energy system. Due to the complexity of the historical building load data, for example the changing energy usage

patterns over the dataset, it was not possible to construct a probabilistic model of the data, and so [Vol](#) couldn't be used. Instead, the historic dataset of building electricity metering data from the Cambridge University estate was used as a case study, and the analysis considered how the forecasting model accuracies would change if more of the data were available when they were trained. A simple linear prediction model was found to achieve similar forecast accuracy to high-complexity, state-of-the-art machine learning models, but was more data efficient and provided better generalisation performance. This was partially due to better computational efficiency of the simple model which meant that it could use a larger context window within the computational budget, and so use more information to support its predictions. These results indicate that in this case, the data used for decision making was more important than the sophistication of the model. Using more than 2 years of hourly resolved load data for model training did not significantly improve the forecasting accuracy, but using less than 1 year led to substantially worse predictions. This indicates that between 1 and 2 years of load data is needed to train an accurate prediction model for building load. However the data used for training is important. Screening training data using change-point analysis to remove low similarity data was able to simultaneously improve data efficiency and prediction accuracy, provided at least 1 year of training data was kept. But it was also found that this data does not necessarily need to come from the target building. Reusing load prediction models trained on data from similar buildings, selected using a data-driven similarity metric, led to prediction errors only 11% higher than models trained on data from the target building. Therefore, the cost of collecting building specific data for model training should be weighed against the benefits that 11% better prediction accuracy provides for building control. To investigate this, the relationship between forecast accuracy and operational performance of [MPC](#) was quantified using synthetic forecasts. The [MPC](#) scheme was much more sensitive to the prediction accuracy of grid electricity price and carbon intensity than building load, suggesting that it is more important to focus expenditure there, for example by purchasing commercial forecasts.

Chapter 5 considered the building load data requirements for the design of a similar district energy system. It quantified the economic benefit of using hourly building energy usage data to support the sizing of solar-battery systems to decarbonise a district of buildings under grid constraints. Probabilistic models of building load profiles for the case study district were constructed using the same dataset of historic metering data from the Cambridge University estate. The uncertainty in building load was found to have a significant impact on the operating cost, causing a $\pm 30\%$ variation in total cost for the energy system designed without monitoring data available. Also, reducing uncertainty in the building load by collecting monitoring data had a significant effect on the optimized energy system design, with total installed capacities of solar, battery, and grid connection varying by

roughly $\pm 20\%$ across the hypothesised measurement values. Therefore, existing methods which stop at either of these two points would conclude that collecting hourly metering data would have a significant effect on system design. However, applying the On-Policy [Vol](#) framework showed that reducing load uncertainty via monitoring would reduce the overall energy system cost by less than 1.5% on average through improved design. This was less than the estimated cost of getting the load data at the time of design, and so using hourly load monitoring data to support the design of the district energy system was found to be not economically worthwhile. Additionally, reducing uncertainty in only the mean load was found to provide most of the available decision support benefit. This means that for building retrofits where the mean building load can be well estimated from historic monthly energy usage data, using hourly metering data would provide little additional benefit for system design, and so should only be used if easily available. With the results from Chapter 4, this indicates that hourly load data from a building is not needed before a solar-battery system can be designed and installed. As the distribution of load profiles from other buildings can be used to make a sufficiently good system design, and a load prediction model can be trained on data from another, similar building to make a sufficiently good control scheme. This is important for building energy system designers and operators, as it allows them to avoid wasteful delays to monitor building loads before installation.

In contrast, Chapter 6 found that reducing uncertainty in the performance of energy storage technologies had a significant impact on energy park design. It investigated the impact of uncertainty in the cost, lifetime, and efficiency of energy storage technologies on the sizing of assets in an energy park, and the benefit of retaining optionality in storage technology choice through the design process. It quantified the improvement in design (reduction in average total cost) achieved when multiple technologies are developed to retain optionality in technology choice, and allow a better storage selection to be made after uncertainty in performance has been reduced. Four candidate storage technologies were considered for use in the park: [Li-ion](#), [NaS](#), [VRFB](#), and [CAES](#). When only a single storage technology was selected for development, updating the system design (wind, solar, and storage sizings) after R&D, when improved information on storage performance is available (i.e. with reduced uncertainty), was found to reduce total costs by 18% on average. This showed that flexibility in procurement contracts for generation and storage is highly valuable to energy park developers. If all four energy storage technologies were developed, so that when the system design was updated any technology could be selected for the final design, the average total cost was found to reduce by 13% compared to the case without optionality. This cost saving was substantially greater than the cost performing R&D for the three additional technologies. This means retaining optionality in storage technology choice is worthwhile, and so developers should hedge their bets against their choice of

storage technology and provide themselves the option to choose the best option once they have better information. However, if two storage technologies were used in the energy park, the average total system cost was 14% lower than the base case without optionality, and providing storage technology optionality was no longer worthwhile. These results were shown to be robust to the level of uncertainty reduction in energy storage performance provided by R&D used in the statistical model, as well as the grid electricity price and carbon emissions data used, and the cost of developing demonstrator storage systems assumed. So for energy park developers, uncertainty in energy storage technology performance is really important for their decision making. Having the option to adjust their design decisions after better information about storage performance is available can significantly reduce costs, in this case by 31%.

The results of this thesis demonstrate the importance of understanding the impact that data has on decision making. By quantifying the effect that data collection and uncertainty reduction have on their decisions, decision makers can either achieve much better outcomes by using information effectively at the right stage in their decision making process, or know when they can make sufficiently good decisions with less data and so avoid wasting resources collecting data that provides little benefit. In the context of decarbonising energy systems, this understanding can unlock substantial cost and carbon savings purely by adjusting the way we make decisions and the data we use to support them. It can also help us identify what we need to know about our energy systems to design and operate them effectively, and determine when we can make good decisions here and now without wasting time waiting for data. The examples in the thesis show that these opportunities exist across all decision making areas in energy systems, from design to operation to management, and across all scales of energy system, from individual buildings to grid-scale energy parks.

The thesis also demonstrated that [Value of Information](#) analysis provides a clear and rigorous methodology for quantifying the benefit that data collection and uncertainty reduction provide to decision making, and showed how it can be used to justify, rationalise, and prioritise expenditure on data collection in energy systems. The methodology is versatile, and can be used to determine whether data collection is worthwhile to support decision making, which and how much data is most economical to collect, and whether it is worthwhile retaining decision optionality until better information is available. It can do so in complex decision making contexts, such as optimisation based design and control. As the value of data collection is often not obvious even to those with expert knowledge of energy systems, it is important to use this systematic approach to understand how data can support decision making in each energy system context.

7.1 Limitations and further work

This thesis presented a series of case studies investigating the impact of particular data collection options on specific decision making tasks in given energy systems. While these case studies aimed to cover a broad range of collected data, decision making contexts, and energy system scales, they are necessarily limited and cannot cover all variations and possibilities that are of interest to energy system decision makers.

The key limitation of this work is the lack of generalisation guarantees for the recommendations about how data should be collected and used to support decision making provided by the numerical results. *Vol* values are only valid under the modelling assumptions used in the calculations, and even innocuous looking changes to the model can lead to very different *Vol* results and conclusions⁶⁵. Similarly, the recommendations about how data should be used to train load forecasting models from Chapter 4 are only evidenced for the university buildings from the Cambridge Estates dataset (see Appendix A). There is no guarantee that this aspect of the dataset is representative of buildings from other universities, for example those in different climates or those with different envelopes and usage patterns, or for other types of buildings, such as residential or commercial buildings. Further, from results on the value of one type of data for supporting a particular decision, we cannot make conclusions about whether other data might be worthwhile collecting, or whether the same data might be worthwhile for supporting a different decision. For example, while Chapter 5 showed that collecting hourly load monitoring data was not worthwhile for informing the sizing of solar-battery systems, we cannot know from this whether improved information about the cost of the solar panels or batteries would have a significant impact on the design decision⁶⁶, or whether collecting hourly load data would be worthwhile to support the sizing of the heat pumps in the buildings.

Significant further research is required to determine if the conclusions from this thesis hold across different energy system contexts, such as location, usage type, and scale. Future work should also use the methodologies outlined in this thesis to investigate other decision tasks and data collection options across the energy systems field, to identify opportunities to improve decision making by providing the right data at the right time, and to avoid wasting time and resources collecting data that provides little insight.

There are also many more aspects of data collection and usage that could be explored. For instance, investigation of how much data should be collected to support decision making. This would require more sophisticated probabilistic models of uncertainty reduction from

⁶⁵ This is discussed in more detail in Section 2.6.

⁶⁶ These uncertainties were not included in the decision model used in Chapter 5.

repeated measurements, but is otherwise straightforward to use within the **Vol** framework. Chapter 4 considered how data availability at the time of model training affects forecasting accuracy. This is a static view of data, but as building energy systems are operated, operational data is available to collect in real-time. Which data should be collected and how this data can be best used to improve building control is an important research question to address. (Raisch, **Langtry**, et al., 2025) has begun to explore the latter question in the context of continual learning for building thermal dynamics.

One of the core ideas of this work is that the purpose of data is to support decisions, and so the value of data should be measured in terms of its ability to improve decision making. Chapter 4 explores how data can be used to develop high accuracy forecasting models, the goal of which is to improve the control of building energy systems, specifically **MPC** in the chapter's case study. The final experiment in the chapter investigated the relationship between forecast accuracy and operation performance of the controller, however as it used relatively simple synthetic forecasts the insight into the control performance with practical forecasts is somewhat limited. Currently in the building **MPC** literature, forecast accuracy and control optimisation are investigated separately. However, to properly understand the data collection requirements of practical building control schemes, these two stages need to be considered together. Chapter 4 provides a starting point for this, but further work is needed to understand where prediction accuracy is important for achieving effective control, and the trade-off between the benefits of accuracy and the costs of achieving it.

From a technical perspective, there are significant research opportunities for improving the computational efficiency of **Vol** calculations. For example, surrogate modelling, importance sampling, or multi-fidelity modelling methods could be used to reduce the number of model evaluations needed to get good estimates of **Vol** values. Identifying appropriate trade-offs between computational cost and **Vol** estimate accuracy would also require better understanding of the accuracy/error of those **Vol** estimates⁶⁷.

Overall, this thesis has demonstrated for a small number of cases the importance of understanding how data and uncertainty reduction affect the decisions we make within energy systems, and has shown how analysing this can improve the decisions we make about data collection. There are many more decisions across energy systems than could be improved by thinking about uncertainty in this way.

⁶⁷ This is discussed further in Section 5.C.

Appendix A. Cambridge University Estates building energy usage dataset

Max Langtry and Ruchi Choudhary (Apr. 2024). *Cambridge University Estates Building Energy Usage Archive*. DOI: [10.5281/zenodo.10955332](https://doi.org/10.5281/zenodo.10955332). URL: <https://github.com/EECi/Cambridge-Estates-Building-Energy-Archive>

This dataset described in this appendix has been released as the above.

The full dataset can be viewed interactively online at <https://eeci.github.io/Cambridge-Estates-Building-Energy-Archive>. Further documentation is available at <https://github.com/EECi/Cambridge-Estates-Building-Energy-Archive>.

The University of Cambridge has been collecting energy usage data from buildings across its estate for several decades. Thanks to this the University Estates Division has accumulated one of the largest, long-duration building energy usage datasets (Kang et al., 2023). It is particularly interesting as it covers buildings of widely varying sizes, ages, and usage types, and includes significant changes in energy usage behaviour, for example after buildings have been retrofit.

The scale of this dataset makes it a valuable resource for building energy research. However previously only small sections of the dataset have been used for research (Pickering and Choudhary, 2019; Ward and Choudhary, 2021; Ward, Choudhary, et al., 2019; Zhuang et al., 2023). To allow researchers to make use of this data resource for large-scale studies, the dataset has been cleaned, formatted, and made publicly available in an anonymised form. This dataset enabled the studies performed in Chapters 4 and 5, and has already been used by other researchers (Choi, S. Lee, Langtry, et al., 2025b).

The dataset contains historic electricity and gas usage data covering the period 2000 to 2023 from 121 buildings in the Cambridge University Estate of varying use types, such as lecture blocks, offices, laboratories, and museums. Due to issues with the building monitoring, the buildings have sometimes sporadic data availability. However in total, there were 1649 years of electricity data and 600 years of gas data with sufficient data quality which have been made available. Alongside this energy usage data, historic weather data, grid electricity price ,and grid carbon intensity data for Cambridge have been provided. The dataset has also been designed to be compatible with the CityLearn framework (Vazquez-Canteli, Dey, et al., 2020; Vázquez-Canteli et al., 2019) to make it easy to simulate building energy system control with real operational data.

The energy usage data was collected by building monitoring systems (smart meters) installed in each building. The electricity usage measurements record the total electrical load at a given metering point in the building, which includes lighting, plug loads, and plant equipment electricity consumption. It is assumed that none of the buildings have heat pumps or AC units, so the electricity usage does not include any contributions from space heating or cooling. The gas usage measurements provide the total gas consumption at the metering points in each building, which is used for both space heating and DHW provision. In the dataset, energy usage data is available at an hourly resolution. The raw data, at a half-hourly resolution, was cleaned by replacing missing values with 0 readings, clipping all values to the range 0 to 10 times the mean of each dataset, and aggregating the data to hourly resolution. Clipping is performed to limit the impact of excessively large observations in the raw data caused by data recording issues in the building monitoring systems.

Hourly weather observation data was collected from the Met Office MIDAS dataset Met Office, 2022. As all buildings in the dataset are closely located, within Cambridge city, the same weather data is used for all buildings. Outdoor dry bulb temperature and relative humidity data are taken from the Bedford measurement station of the MIDAS dataset, the closest location with complete data availability over the period, which provides pre-processed historic weather measurements. Data for direct and diffuse solar irradiance at ground level, and solar generation potential in central Cambridge were collected from the [renewables.ninja](#) reanalysis model (Pfenninger and Staffell, 2016; Staffell and Pfenninger, 2016b), which estimates historic solar irradiance values from historic global weather observations. Hourly dynamic electricity pricing tariff data and grid electricity carbon intensity data were collected from Energy Stats UK (Energy Stats UK, 2023) and the National Grid ESO Data Portal (National Grid ESO, 2020), respectively. Temporal variables are also provided: month, hour, day of the week, and daylight savings status.

Table A.1 summarises the variables provided in the dataset. Further detail on the data sourcing and processing is available at <https://github.com/EECi/Cambridge-Estates-Building-Energy-Archive>. Fig. A.1 summarises the data availability across buildings in the dataset.

Variable	Units	Data availability	Data source
<i>Building specific variables</i>			
Electrical load	kWh	Building dependent	Cambridge estates building meters
Gas load	kWh		
<i>Location related variables</i>			
Grid electricity price	£/kWh	2019 to 2023	Energy Stats UK (Energy Stats UK, 2023)
Grid electricity carbon intensity	kgCO ₂ /kWh	2009 to 2023	National Grid ESO data portal (National Grid ESO, 2020)
Solar generation potential	W/kWp		<i>renewables.ninja</i> reanalysis model
Direct solar irradiance	W/m ²	All	(Pfenninger and Staffell, 2016) (Staffell and Pfenninger, 2016b)
Diffuse solar irradiance	W/m ²		
Outdoor temperature	°C	2000 to 2022	Met Office MIDAS dataset
Relative humidity	%		(Met Office, 2022)
Time information	—	All	—

Table A.1 Summary of variables available in dataset.

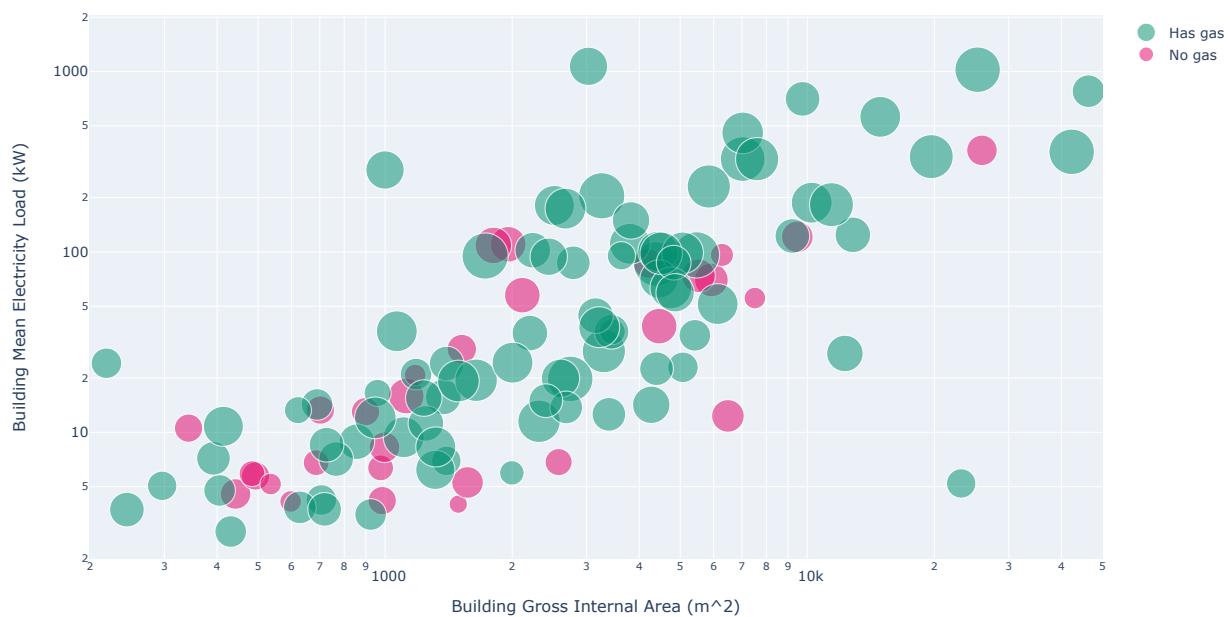


Fig. A.1 Summary of data availability across buildings in dataset.

Interactive version of plot available online [here](#).

References

- Acar, Erdem and Kiran Solanki (Sept. 2009). "System Reliability Based Vehicle Design for Crash-worthiness and Effects of Various Uncertainty Reduction Measures". In: *Structural and Multidisciplinary Optimization* 39.3, pp. 311–325. ISSN: 1615-1488. DOI: [10.1007/s00158-008-0327-3](https://doi.org/10.1007/s00158-008-0327-3). URL: <https://doi.org/10.1007/s00158-008-0327-3>.
- Acha, Salvador, Guilherme Vieira, Max Bird, and Nilay Shah (Sept. 2022). "Modelling UK Electricity Regional Costs for Commercial Buildings". In: *Energy and Buildings* 271, p. 112301. ISSN: 0378-7788. DOI: [10.1016/j.enbuild.2022.112301](https://doi.org/10.1016/j.enbuild.2022.112301). URL: <https://www.sciencedirect.com/science/article/pii/S0378778822004728>.
- Al Dakheel, Joud, Claudio Del Pero, Niccolò Aste, and Fabrizio Leonforte (Oct. 2020). "Smart Buildings Features and Key Performance Indicators: A Review". In: *Sustainable Cities and Society* 61, p. 102328. ISSN: 2210-6707. DOI: [10.1016/j.scs.2020.102328](https://doi.org/10.1016/j.scs.2020.102328). URL: <https://www.sciencedirect.com/science/article/pii/S2210670720305497>.
- Alahakoon, Damminda and Xinghuo Yu (Feb. 2016). "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey". In: *IEEE Transactions on Industrial Informatics* 12.1, pp. 425–436. ISSN: 1941-0050. DOI: [10.1109/TII.2015.2414355](https://doi.org/10.1109/TII.2015.2414355). URL: <https://ieeexplore.ieee.org/abstract/document/7063262>.
- Alonso-Travesset, Àlex, Helena Martín, Sergio Coronas, and Jordi de la Hoz (Jan. 2022). "Optimization Models under Uncertainty in Distributed Generation Systems: A Review". In: *Energies* 15.5, p. 1932. ISSN: 1996-1073. DOI: [10.3390/en15051932](https://doi.org/10.3390/en15051932). URL: <https://www.mdpi.com/1996-1073/15/5/1932>.
- Amini Toosi, Hashem, Monica Lavagna, Fabrizio Leonforte, Claudio Del Pero, and Niccolò Aste (Dec. 2022). "Building Decarbonization: Assessing the Potential of Building-Integrated Photovoltaics and Thermal Energy Storage Systems". In: *Energy Reports. Selected Papers from 2022 7th International Conference on Advances on Clean Energy Research* 8, pp. 574–581. ISSN: 2352-4847. DOI: [10.1016/j.egyr.2022.10.322](https://doi.org/10.1016/j.egyr.2022.10.322). URL: <https://www.sciencedirect.com/science/article/pii/S2352484722022570>.
- Anderson, Hope C., Abdullah Al Hadi, Evan S. Jones, and Dan M. Ionel (Sept. 2021). "Power Factor and Reactive Power in US Residences – Survey and EnergyPlus Modeling". In: *2021 10th International Conference on Renewable Energy Research and Application (ICRERA)*. Istanbul, Turkey: IEEE, pp. 418–422. ISBN: 978-1-66544-524-5. DOI: [10.1109/ICRERA52334.2021.9598561](https://doi.org/10.1109/ICRERA52334.2021.9598561). URL: <https://ieeexplore.ieee.org/document/9598561/>.
- Aresti, Lazaros, Paul Christodoulides, and Georgios Florides (Sept. 2018). "A Review of the Design Aspects of Ground Heat Exchangers". In: *Renewable and Sustainable Energy Reviews* 92, pp. 757–773. ISSN: 1364-0321. DOI: [10.1016/j.rser.2018.04.053](https://doi.org/10.1016/j.rser.2018.04.053). URL: <https://www.sciencedirect.com/science/article/pii/S1364032118302727>.
- Arévalo, Paul and Francisco Jurado (Feb. 2021). "Performance Analysis of a PV/HKT/WT/DG Hybrid Autonomous Grid". In: *Electrical Engineering* 103.1, pp. 227–244. ISSN: 1432-0487. DOI: [10.1007/s00202-020-01065-9](https://doi.org/10.1007/s00202-020-01065-9). URL: <https://doi.org/10.1007/s00202-020-01065-9>.

- ASHRAE (Sept. 2024). *Energy Calculations: ASHRAE Technical Committee 4.7*. URL: <https://web.archive.org/web/20240919013441/https://tpc.ashrae.org/CustomPages?idx=5f0b34a1-4e42-45cd-b925-0f711f5d79aa>.
- Auchincloss, Murray (Feb. 2024). “2023 : A Year of Delivery – Financial Summary”. In: *BP*. URL: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/investors/bp-fourth-quarter-2023-results.pdf>.
- Avriel, M. and A. C. Williams (1970). “The Value of Information and Stochastic Programming”. In: *Operations Research* 18.5, pp. 947–954. ISSN: 0030-364X. JSTOR: 169369. URL: <https://www.jstor.org/stable/169369>.
- Bakke, Ida, Stein-Erik Fleten, Lars Ivar Hagfors, et al. (July 2016). “Investment in Electric Energy Storage under Uncertainty: A Real Options Approach”. In: *Computational Management Science* 13.3, pp. 483–500. ISSN: 1619-6988. doi: 10.1007/s10287-016-0256-3. URL: <https://doi.org/10.1007/s10287-016-0256-3>.
- Balali, Yasaman, Adrian Chong, Andrew Busch, and Steven O’Keefe (Sept. 2023). “Energy Modelling and Control of Building Heating and Cooling Systems with Data-Driven and Hybrid Models—A Review”. In: *Renewable and Sustainable Energy Reviews* 183, p. 113496. ISSN: 1364-0321. doi: 10.1016/j.rser.2023.113496. URL: <https://www.sciencedirect.com/science/article/pii/S1364032123003532>.
- Barber, Kyle A. and Moncef Krarti (May 2022). “A Review of Optimization Based Tools for Design and Control of Building Energy Systems”. In: *Renewable and Sustainable Energy Reviews* 160, p. 112359. ISSN: 1364-0321. doi: 10.1016/j.rser.2022.112359. URL: <https://www.sciencedirect.com/science/article/pii/S1364032122002696>.
- Bartolucci, Lorenzo, Stefano Cordiner, Vincenzo Mulone, and Marina Santarelli (Apr. 2019). “Hybrid Renewable Energy Systems: Influence of Short Term Forecasting on Model Predictive Control Performance”. In: *Energy* 172, pp. 997–1004. ISSN: 0360-5442. doi: 10.1016/j.energy.2019.01.104. URL: <https://www.sciencedirect.com/science/article/pii/S0360544219301045>.
- Bedford, Norton M and Mohamed Onsi (1966). “Measuring the Value of Information - An Information Theory Approach”. In: *Management Services: A Magazine of Planning, Systems, and Controls* 3.1. URL: <https://egrove.olemiss.edu/mgmtservices/vol3/iss1/3/>.
- BEIS (Mar. 2023). *Powering Up Britain*. Tech. rep. HM Government. URL: <https://assets.publishing.service.gov.uk/media/642468ff2fa8480013ec0f39/powering-up-britain-joint-overview.pdf>.
- Beitner, Jan (Jan. 2024). *Jdb78/Pytorch-Forecasting*. URL: <https://github.com/jdb78/pytorch-forecasting>.
- Bistline, John E. (Sept. 2015). “Electric Sector Capacity Planning under Uncertainty: Climate Policy and Natural Gas in the US”. In: *Energy Economics* 51, pp. 236–251. ISSN: 0140-9883. doi: 10.1016/j.eneco.2015.07.008. URL: <https://www.sciencedirect.com/science/article/pii/S0140988315002157>.
- Blonsky, Michael, Adarsh Nagarajan, Shibani Ghosh, et al. (Dec. 2019). “Potential Impacts of Transportation and Building Electrification on the Grid: A Review of Electrification Projections and Their Effects on Grid Infrastructure, Operation, and Planning”. In: *Current Sustainable/Renewable Energy Reports* 6.4, pp. 169–176. ISSN: 2196-3010. doi: 10.1007/s40518-019-00140-5. URL: <https://doi.org/10.1007/s40518-019-00140-5>.

- Bloomberg (May 2024). *EU ETS Market Outlook 1H 2024: Prices Valley Before Rally* | BloombergNEF. URL: <https://web.archive.org/web/20240523100529/https://about.bnef.com/blog/eu-ets-market-outlook-1h-2024-prices-valley-before-rally/>.
- Bonnans, Joseph-Frédéric, Jean Charles Gilbert, Claude Lemarechal, and Claudia A. Sagastizábal (Oct. 2006). *Numerical Optimization: Theoretical and Practical Aspects*. Springer Science & Business Media. ISBN: 978-3-540-35447-5.
- Bortolini, Rafaela, Raul Rodrigues, Hamidreza Alavi, Luisa Felix Dalla Vecchia, and Núria Forcada (Jan. 2022). “Digital Twins’ Applications for Building Energy Efficiency: A Review”. In: *Energies* 15.19, p. 7002. ISSN: 1996-1073. doi: [10.3390/en15197002](https://doi.org/10.3390/en15197002). URL: <https://www.mdpi.com/1996-1073/15/19/7002>.
- Boyd, Stephen P. and Lieven Vandenberghe (Mar. 2004). *Convex Optimization*. Cambridge University Press. ISBN: 978-0-521-83378-3.
- BP (Dec. 2022). *Green hydrogen project H2-Fifty selected for IPCEI grant funding* | Nieuws | Home. URL: https://www.bp.com/nl_nl/netherlands/home/nieuws/persberichten/Green_hydrogen_project_H2-Fifty_selected_for_IPCEI_grant_funding.html.
- Bravo Dias, João, Guilherme Carrilho da Graça, and Pedro M. M. Soares (Jan. 2020). “Comparison of Methodologies for Generation of Future Weather Data for Building Thermal Energy Simulation”. In: *Energy and Buildings* 206, p. 109556. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2019.109556](https://doi.org/10.1016/j.enbuild.2019.109556). URL: <https://www.sciencedirect.com/science/article/pii/S037877881931552X>.
- Brown, Tom, Jonas Hörsch, and David Schlachtberger (Jan. 2018). “PyPSA: Python for Power System Analysis”. In: *Journal of Open Research Software* 6.1, p. 4. ISSN: 2049-9647. doi: [10.5334/jors.188](https://doi.org/10.5334/jors.188). URL: <http://openresearchsoftware.metajnl.com/articles/10.5334/jors.188/>.
- Brown, Tom, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner (Oct. 2018). “Synergies of Sector Coupling and Transmission Reinforcement in a Cost-Optimised, Highly Renewable European Energy System”. In: *Energy* 160, pp. 720–739. ISSN: 0360-5442. doi: [10.1016/j.energy.2018.06.222](https://doi.org/10.1016/j.energy.2018.06.222). URL: <https://www.sciencedirect.com/science/article/pii/S036054421831288X>.
- Büning, Felix, Benjamin Huber, Adrian Schalbetter, et al. (Mar. 2022). “Physics-Informed Linear Regression Is Competitive with Two Machine Learning Methods in Residential Building MPC”. In: *Applied Energy* 310, p. 118491. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2021.118491](https://doi.org/10.1016/j.apenergy.2021.118491). URL: <https://www.sciencedirect.com/science/article/pii/S0306261921017098>.
- Büning, Felix, Corentin Pfister, Ahmed Aboudonia, Philipp Heer, and John Lygeros (2021). “Comparing Machine Learning Based Methods to Standard Regression Methods for MPC on a Virtual Testbed”. In: *Building Simulation 2021*. Vol. 17. Building Simulation. IBPSA, pp. 127–134. ISBN: 978-1-77505-202-9. doi: [10.26868/25222708.2021.30346](https://doi.org/10.26868/25222708.2021.30346). URL: https://publications.ibpsa.org/conference/paper/?id=bs2021_30346.
- Busby, Jon (June 2018). “A Modelling Study of the Variation of Thermal Conductivity of the English Chalk”. In: *Quarterly Journal of Engineering Geology and Hydrogeology* 51.4, pp. 417–423. ISSN: 1470-9236. doi: [10.1144/qjegh2017-127](https://doi.org/10.1144/qjegh2017-127). URL: <https://doi.org/10.1144/qjegh2017-127>.
- (2011). “Provision of Thermal Properties Data for Ground Collector Loop Design”. In: URL: https://mail.gshp.org.uk/pdf/CambridgeSeminar2011/JonBusby_BGS_Thermal_Conductivity.pdf.
- Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, et al. (2017). “Stan: A Probabilistic Programming Language”. In: *Journal of statistical software* 76, p. 1. ISSN: 1548-7660. doi: [10.1863/7/jss.v076.i01](https://doi.org/10.1863/7/jss.v076.i01). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9788645/>.

- Chadly, Assia, Rahul Rajeevkumar Urs, Max Wei, Maher Maalouf, and Ahmad Mayyas (July 2023). "Techno-Economic Assessment of Energy Storage Systems in Green Buildings While Considering Demand Uncertainty". In: *Energy and Buildings* 291, p. 113130. ISSN: 0378-7788. DOI: [10.1016/j.enbuild.2023.113130](https://doi.org/10.1016/j.enbuild.2023.113130). URL: <https://www.sciencedirect.com/science/article/pii/S0378778823003602>.
- Challu, Cristian, Kin G. Olivares, Boris N. Oreshkin, et al. (June 2023). "NHITS: Neural Hierarchical Interpolation for Time Series Forecasting". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 37.6, pp. 6989–6997. ISSN: 2374-3468. DOI: [10.1609/aaai.v37i6.25854](https://doi.org/10.1609/aaai.v37i6.25854). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/25854>.
- Charitopoulos, Vassilis M., Mathilde Fajard, Chi Kong Chyong, and David M. Reiner (Nov. 2023). "The Impact of 100% Electrification of Domestic Heat in Great Britain". In: *iScience* 26.11, p. 108239. ISSN: 2589-0042. DOI: [10.1016/j.isci.2023.108239](https://doi.org/10.1016/j.isci.2023.108239). URL: <https://www.sciencedirect.com/science/article/pii/S2589004223023167>.
- Chinaris, Periklis P., Georgios N. Psarros, and Stavros A. Papathanassiou (Feb. 2025). "Hybridization of Wind Farms with Co-Located PV and Storage Installations". In: *Renewable Energy* 240, p. 122057. ISSN: 0960-1481. DOI: [10.1016/j.renene.2024.122057](https://doi.org/10.1016/j.renene.2024.122057). URL: <https://www.sciencedirect.com/science/article/pii/S0960148124021256>.
- Choi, Wonjun, Ruchi Choudhary, and Ryozo Ooka (May 2021). "Development of Chiller-Attached Apparatus for Accurate Initial Ground Temperature Measurement: Insights from Global Sensitivity Analysis of Thermal Response Tests". In: *Energy and Buildings* 238, p. 110841. ISSN: 0378-7788. DOI: [10.1016/j.enbuild.2021.110841](https://doi.org/10.1016/j.enbuild.2021.110841). URL: <https://www.sciencedirect.com/science/article/pii/S0378778821001250>.
- Choi, Wonjun and Sangwon Lee (June 2023). "Performance Evaluation of Deep Learning Architectures for Load and Temperature Forecasting under Dataset Size Constraints and Seasonality". In: *Energy and Buildings* 288, p. 113027. ISSN: 0378-7788. DOI: [10.1016/j.enbuild.2023.113027](https://doi.org/10.1016/j.enbuild.2023.113027). URL: <https://www.sciencedirect.com/science/article/pii/S0378778823002578>.
- Choi, Wonjun, Sangwon Lee, **Max Langtry**, and Ruchi Choudhary (Sept. 2025a). *Privacy-Preserving Transfer Learning Framework for Building Energy Forecasting with Fully Anonymized Data*. (Working paper).
- (Feb. 2025b). "Representation Vector-Based Time Series Similarity Analysis Using Unsupervised Contrastive Learning Time Series Encoder". In: *Korean Journal of Air-Conditioning and Refrigeration Engineering* 37.2, pp. 72–81. ISSN: 1229-6422. DOI: [10.6110/KJACR.2025.37.2.72](https://doi.org/10.6110/KJACR.2025.37.2.72). URL: <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE12036797>.
- Chong, Adrian and Kathrin Menberg (Sept. 2018). "Guidelines for the Bayesian Calibration of Building Energy Models". In: *Energy and Buildings* 174, pp. 527–547. ISSN: 0378-7788. DOI: [10.1016/j.enbuild.2018.06.028](https://doi.org/10.1016/j.enbuild.2018.06.028). URL: <https://www.sciencedirect.com/science/article/pii/S037878818307539>.
- Chong, Adrian, Weili Xu, and Khee Poh Lam (Dec. 2015). "Uncertainty Analysis in Building Energy Simulation. A Practical Approach". In: *2015 Building Simulation Conference*. DOI: [10.26868/2522708.2015.2131](https://doi.org/10.26868/2522708.2015.2131). URL: https://publications.ibpsa.org/conference/paper/?id=bs2015_2131.
- Chou, Jui-Sheng and Duc-Son Tran (Dec. 2018). "Forecasting Energy Consumption Time Series Using Machine Learning Techniques Based on Usage Patterns of Residential Householders". In: *Energy* 165, pp. 709–726. ISSN: 0360-5442. DOI: [10.1016/j.energy.2018.09.144](https://doi.org/10.1016/j.energy.2018.09.144). URL: <https://www.sciencedirect.com/science/article/pii/S0360544218319145>.

- Clements, Derek (2020). *The Impact of Sub-Metering Requirements on Building Electrical Systems Design*. Tech. rep. Kansas State University. URL: <https://krex.k-state.edu/server/api/core/bitstreams/6826e835-ef68-4f5a-8756-297697591027/content>.
- CNZ (Jan. 2025). *Net Zero Building Metrics: Smart Homes for the Future Energy System*. Tech. rep. Centre for Net Zero. URL: <https://www.centrefornetzero.org/papers/net-zero-building-metrics-smart-homes-for-the-future-energy-system>.
- Commission, JRC Euproean (Sept. 2017). *JRC Photovoltaic Geographical Information System (PVGIS)*. URL: https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis/pvgis-releases/pvgis-5_en.
- Committee on Climate Change (Dec. 2020). *The Sixth Carbon Budget The UK's Path to Net Zero*. Tech. rep. Committee on Climate Change. URL: <https://www.theccc.org.uk/wp-content/uploads/2020/12/The-Sixth-Carbon-Budget-The-UKs-path-to-Net-Zero.pdf>.
- Connolly, D., H. Lund, B. V. Mathiesen, and M. Leahy (Apr. 2010). "A Review of Computer Tools for Analysing the Integration of Renewable Energy into Various Energy Systems". In: *Applied Energy* 87.4, pp. 1059–1082. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2009.09.026](https://doi.org/10.1016/j.apenergy.2009.09.026). URL: <https://www.sciencedirect.com/science/article/pii/S0306261909004188>.
- Conrad, Jon M. (June 1980). "Quasi-Option Value and the Expected Value of Information". In: *The Quarterly Journal of Economics* 94.4, pp. 813–820. ISSN: 0033-5533. doi: [10.2307/1885672](https://doi.org/10.2307/1885672). URL: <https://doi.org/10.2307/1885672>.
- Coppitters, Diederik, Ward De Paepe, and Francesco Contino (Aug. 2021). "Robust Design Optimization of a Photovoltaic-Battery-Heat Pump System with Thermal Storage under Aleatory and Epistemic Uncertainty". In: *Energy* 229, p. 120692. ISSN: 0360-5442. doi: [10.1016/j.energy.2021.120692](https://doi.org/10.1016/j.energy.2021.120692). URL: <https://www.sciencedirect.com/science/article/pii/S0360544221009403>.
- Crawley, Drury and Linda Lawrie (Dec. 2015). *Rethinking the TMY: Is the 'Typical' Meteorological Year Best for Building Performance Simulation?* doi: [10.26868/25222708.2015.2707](https://doi.org/10.26868/25222708.2015.2707).
- Dai, Ting-Yu, Ayegül Demir Dilsiz, Dev Niyogi, and Zoltan Nagy (Nov. 2023). "A Comparison of Different Deep Learning Model Architectures and Training Strategy for Urban Energy Modeling". In: *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. Istanbul Turkey: ACM, pp. 316–317. ISBN: 9798400702303. doi: [10.1145/3600100.3626277](https://doi.org/10.1145/3600100.3626277). URL: <https://dl.acm.org/doi/10.1145/3600100.3626277>.
- Dai, Ting-Yu, Dev Niyogi, and Zoltan Nagy (Dec. 2023). *CityTFT: Temporal Fusion Transformer for Urban Building Energy Modeling*. arXiv: [2312.02375 \[cs, stat\]](https://arxiv.org/abs/2312.02375). URL: [http://arxiv.org/abs/2312.02375](https://arxiv.org/abs/2312.02375).
- Daikin (Apr. 2022). *Daikin UK Price Book*. Tech. rep. URL: https://sbs.co.uk/file/Supplier/Daikin_Price_List.pdf.
- Dalla Santa, Giorgia, Antonio Galgaro, Raffaele Sassi, et al. (May 2020). "An Updated Ground Thermal Properties Database for GSHP Applications". In: *Geothermics* 85, p. 101758. ISSN: 0375-6505. doi: [10.1016/j.geothermics.2019.101758](https://doi.org/10.1016/j.geothermics.2019.101758). URL: <https://www.sciencedirect.com/science/article/pii/S0375650519301944>.
- Dantzig, George B. (Apr. 1955). "Linear Programming under Uncertainty". In: *Management Science* 1.3-4, pp. 197–206. ISSN: 0025-1909. doi: [10.1287/mnsc.1.3-4.197](https://doi.org/10.1287/mnsc.1.3-4.197). URL: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1.3-4.197>.

- Dantzig, George B. (1956). "Recent Advances in Linear Programming". In: *Management Science* 2.2, pp. 131–144. ISSN: 0025-1909. JSTOR: 2627490. URL: <https://www.jstor.org/stable/2627490>.
- de Oliveira, P. M., L. C. C. Mesquita, S. Gkantonas, A. Giusti, and E. Mastorakos (Jan. 2021). "Evolution of Spray and Aerosol from Respiratory Releases: Theoretical Estimates for Insight on Viral Transmission". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477.2245, p. 20200584. doi: 10.1098/rspa.2020.0584. URL: <https://royalsocietypublishing.org/doi/full/10.1098/rspa.2020.0584>.
- DeCarolis, Joseph, Hannah Daly, Paul Dodds, et al. (May 2017). "Formalizing Best Practice for Energy System Optimization Modelling". In: *Applied Energy* 194, pp. 184–198. ISSN: 0306-2619. doi: 10.1016/j.apenergy.2017.03.001. URL: <https://www.sciencedirect.com/science/article/pii/S0306261917302192>.
- Deng, Zhipeng, Xuezheng Wang, Zixin Jiang, et al. (May 2023). "Evaluation of Deploying Data-Driven Predictive Controls in Buildings on a Large Scale for Greenhouse Gas Emission Reduction". In: *Energy* 270, p. 126934. ISSN: 0360-5442. doi: 10.1016/j.energy.2023.126934. URL: <https://www.sciencedirect.com/science/article/pii/S0360544223003286>.
- Department for Business and Trade (Oct. 2023). *UK Battery Strategy*. Tech. rep. HM Government. URL: <https://assets.publishing.service.gov.uk/media/656ef4871104cf000dfa74f3/uk-battery-strategy.pdf>.
- Deroubaix, Adrien, Inga Labuhn, Marie Camredon, et al. (Aug. 2021). "Large Uncertainties in Trends of Energy Demand for Heating and Cooling under Climate Change". In: *Nature Communications* 12.1, p. 5197. ISSN: 2041-1723. doi: 10.1038/s41467-021-25504-8. URL: <https://www.nature.com/articles/s41467-021-25504-8>.
- DESNZ (Dec. 2023a). *Annual Domestic Energy Bills*. URL: <https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics>.
- (July 2024a). *Digest of UK Energy Statistics (DUKES): Electricity*. URL: <https://www.gov.uk/government/statistics/electricity-chapter-5-digest-of-united-kingdom-energy-statistics-dukes>.
- (Nov. 2023b). *Electricity Generation Costs 2023*. Tech. rep. Department for Energy Security & Net Zero. URL: <https://assets.publishing.service.gov.uk/media/6556027d046ed400148b99fe/electricity-generation-costs-2023.pdf>.
- (Apr. 2023c). *Longer Duration Energy Storage Demonstration Programme, Stream 1 Phase 1 Projects*. URL: <https://www.gov.uk/government/publications/longer-duration-energy-storage-demonstration-programme-successful-projects/longer-duration-energy-storage-demonstration-programme-stream-1-phase-1-details-of-successful-projects>.
- (Apr. 2023d). *Longer Duration Energy Storage Demonstration Programme, Stream 1 Phase 2 Project*. URL: <https://www.gov.uk/government/publications/longer-duration-energy-storage-demonstration-programme-successful-projects/longer-duration-energy-storage-demonstration-programme-stream-1-phase-2-details-of-successful-project>.
- (Apr. 2023e). *Longer Duration Energy Storage Demonstration Programme, Stream 2 Phase 2 Projects*. URL: <https://www.gov.uk/government/publications/longer-duration-energy-storage-demonstration-programme-successful-projects/longer-duration-energy-storage-demonstration-programme-stream-2-phase-2-details-of-successful-projects>.
- (June 2024b). *National Energy Efficiency Data-Framework (NEED): Summary of Analysis, Great Britain, 2024*. Tech. rep. Department for Energy Security & Net Zero. URL: <https://assets.publishing.service.gov.uk/media/66e0203ad65d5c23df086710/NEED-report-june-2024.pdf>.

- (Aug. 2024c). *The Non-Domestic National Energy Efficiency Data-Framework: 2024 (England and Wales)*. Tech. rep. Department for Energy Security & Net Zero. URL: <https://assets.publishing.service.gov.uk/media/66b4dfe6ab418ab055593520/ND-NEED-2024-report.pdf>.
 - (Oct. 2022). *UK ETS: Carbon Prices for Use in Civil Penalties, 2023*. URL: <https://www.gov.uk/government/publications/determinations-of-the-uk-ets-carbon-price/uk-ets-carbon-prices-for-use-in-civil-penalties-2023>.
 - (June 2023f). *Update on the Rollout of Smart Meters*. Tech. rep. Department for Energy Security & Net Zero. URL: <https://www.nao.org.uk/wp-content/uploads/2023/06/update-on-the-rollout-of-smart-meters.pdf>.
- Di Francesco, Domenic (Feb. 2023). *Guidance on the Use of Probabilistic Methods for Identifying Data Requirements*. URL: https://github.com/DomDF/DCE_guidance.
- Di Francesco, Domenic, Marios Chryssanthopoulos, Michael Havbro Faber, and Ujjwal Bharadwaj (2021). “Decision-Theoretic Inspection Planning Using Imperfect and Incomplete Data”. In: *Data-Centric Engineering 2*, e18. ISSN: 2632-6736. doi: [10.1017/dce.2021.18](https://doi.org/10.1017/dce.2021.18). URL: https://www.cambridge.org/core/product/identifier/S2632673621000186/type/journal_article.
- Di Francesco, Domenic, **Max Langtry**, Andrew B. Duncan, and Chris Dent (Sept. 2023). *System Effects in Identifying Risk-Optimal Data Requirements for Digital Twins of Structures*. doi: [10.48550/arXiv.2309.07695](https://doi.org/10.48550/arXiv.2309.07695). arXiv: [2309.07695 \[stat\]](https://arxiv.org/abs/2309.07695). URL: [http://arxiv.org/abs/2309.07695](https://arxiv.org/abs/2309.07695).
- Drgoňa, Ján, Javier Arroyo, Iago Cupeiro Figueroa, et al. (Jan. 2020). “All You Need to Know about Model Predictive Control for Buildings”. In: *Annual Reviews in Control* 50, pp. 190–232. ISSN: 1367-5788. doi: [10.1016/j.arcontrol.2020.09.001](https://doi.org/10.1016/j.arcontrol.2020.09.001). URL: <https://www.sciencedirect.com/science/article/pii/S1367578820300584>.
- E.ON Energy (Jan. 2024). *Solar Panel Cost*. URL: <https://web.archive.org/web/20240128072106/https://www.eonenergy.com/solar-panels/cost.html>.
- Economics, Frontier (Dec. 2022). *Network Tariffs for Energy Intensive Industry: Evaluation of Rationale for Exemptions*. Tech. rep. Frontier Economics. URL: <https://assets.publishing.service.gov.uk/media/649be812f90109001281891c/annex-b-frontier-economics-report-on-network-tariffs-for-energy-intensive-industries.pdf>.
- EEA (Feb. 2023). *Decarbonising Heating and Cooling — a Climate Imperative*. Briefing. European Environment Agency. URL: <https://www.eea.europa.eu/publications/decarbonisation-heating-and-cooling>.
- EEPowers (Sept. 2022). *Power Factor: Determining How Much Electricity Your Power System Consumes - Technical Articles*. URL: <https://eepower.com/technical-articles/power-factor-determining-how-much-electricity-your-power-system-consumes/>.
- Egging-Bratseth, Ruud, Hanne Kauko, Brage Rugstad Knudsen, et al. (Mar. 2021). “Seasonal Storage and Demand Side Management in District Heating Systems with Demand Uncertainty”. In: *Applied Energy* 285, p. 116392. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2020.116392](https://doi.org/10.1016/j.apenergy.2020.116392). URL: <https://www.sciencedirect.com/science/article/pii/S0306261920317645>.
- Energy Stats UK (July 2023). *Historical Pricing Data – Octopus Agile Eastern England*. URL: <https://energy-stats.uk/download-historical-pricing-data/>.
- Enríquez, Ricardo, María José Jiménez, and M del Rosario Heras (June 2016). “Solar Forecasting Requirements for Buildings MPC”. In: *Energy Procedia*. Proceedings of the 4th International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2015) 91, pp. 1024–

1032. ISSN: 1876-6102. doi: 10.1016/j.egypro.2016.06.271. URL: <https://www.sciencedirect.com/science/article/pii/S187661021630371X>.
- Erfani, Arash, Tohid Jafarinejad, Staf Roels, and Dirk Saelens (2023). "Linking Dataset Quality and MPC in Buildings: Impact of Temporal Resolution". In: *13th Nordic Symposium on Building Physics (NSB 2023)*. URL: <https://lirias.kuleuven.be/retrieve/729236>.
- Erfani, Arash, Xingji Yu, Tuule Mall Kull, et al. (Sept. 2021). "Analysis of the Impact of Predictive Models on the Quality of the Model Predictive Control for an Experimental Building". In: *Proceedings of Building Simulation 2021*. International Building Performance Simulation Association (IBPSA). URL: <https://lirias.kuleuven.be/3553319>.
- Esnaashary Esfahani, Mansour, Chris Rausch, Carl Haas, and Bryan T. Adey (Sept. 2020). "Prioritizing Preproject Planning Activities Using Value of Information Analysis". In: *Journal of Management in Engineering* 36.5, p. 04020064. ISSN: 1943-5479. doi: 10.1061/(ASCE)ME.1943-5479.0000822. URL: <https://ascelibrary.org/doi/10.1061/%28ASCE%29ME.1943-5479.0000822>.
- ESO, National Grid (Mar. 2024). *Beyond 2030 – A National Blueprint for a Decarbonised Electricity System in Great Britain*. Tech. rep. URL: <https://www.neso.energy/document/315516/download>.
- (July 2023). *Future Energy Scenarios*. Tech. rep. London, UK: National Grid Electricity System Operator. URL: <https://www.nationalgrideso.com/future-energy/future-energy-scenarios-fes>.
- European Union (June 2021). *Regulation (EU) 2021/1119 of the European Parliament and of the Council of 30 June 2021 Establishing the Framework for Achieving Climate Neutrality and Amending Regulations (EC) No 401/2009 and (EU) 2018/1999 ('European Climate Law')*. URL: <http://data.europa.eu/eli/reg/2021/1119/oj/eng>.
- Falcon, William and The PyTorch Lightning team (Mar. 2019). *PyTorch Lightning*. doi: 10.5281/zenodo.3828935. URL: <https://github.com/Lightning-AI/lightning>.
- Fan, Cheng, Jiayuan Wang, Wenjie Gang, and Shenghan Li (Feb. 2019). "Assessment of Deep Recurrent Neural Network-Based Strategies for Short-Term Building Energy Predictions". In: *Applied Energy* 236, pp. 700–710. ISSN: 0306-2619. doi: 10.1016/j.apenergy.2018.12.004. URL: <https://www.sciencedirect.com/science/article/pii/S0306261918318221>.
- Fan, Fulin, Giorgio Zorzi, David Campos-Gaona, et al. (Jan. 2021). "Sizing and Coordination Strategies of Battery Energy Storage System Co-Located with Wind Farm: The UK Perspective". In: *Energies* 14.5, p. 1439. ISSN: 1996-1073. doi: 10.3390/en14051439. URL: <https://www.mdpi.com/1996-1073/14/5/1439>.
- Farrar, Donald E. and Robert R. Glauber (1967). "Multicollinearity in Regression Analysis: The Problem Revisited". In: *The Review of Economics and Statistics* 49.1, pp. 92–107. ISSN: 0034-6535. doi: 10.2307/1937887. JSTOR: 1937887. URL: <https://www.jstor.org/stable/1937887>.
- Farrokhifar, Meisam, Hamidreza Bahmani, Behdad Faridpak, et al. (Dec. 2021). "Model Predictive Control for Demand Side Management in Buildings: A Survey". In: *Sustainable Cities and Society* 75, p. 103381. ISSN: 2210-6707. doi: 10.1016/j.scs.2021.103381. URL: <https://www.sciencedirect.com/science/article/pii/S2210670721006545>.
- Feltham, Gerald A. (1968). "The Value of Information". In: *The Accounting Review* 43.4, pp. 684–696. ISSN: 0001-4826. JSTOR: 243630. URL: <https://www.jstor.org/stable/243630>.
- Fennell, Pamela J., Shima Ebrahimigharehbaghi, Érika Mata, et al. (Feb. 2025). "Urban Building Energy Models: How Can We Improve the Treatment of Uncertainty for Energy Policy Decision-Making?" In: *Environmental Research Communications* 7.2, p. 022002. ISSN: 2515-7620. doi:

- 10.1088/2515-7620/ad9438. URL: <https://iopscience.iop.org/article/10.1088/2515-7620/ad9438/meta>.
- Feydy, Jean (Jan. 2024). *Jeanfeydy/Geomloss*. URL: <https://github.com/jeanfeydy/geomloss>.
- Fiorentini, Massimo, Philipp Heer, and Luca Baldini (Jan. 2023). "Design Optimization of a District Heating and Cooling System with a Borehole Seasonal Thermal Energy Storage". In: *Energy* 262, p. 125464. ISSN: 0360-5442. DOI: 10.1016/j.energy.2022.125464. URL: <https://www.sciencedirect.com/science/article/pii/S0360544222023465>.
- Fodstad, Marte, Pedro Crespo del Granado, Lars Hellemo, et al. (May 2022). "Next Frontiers in Energy System Modelling: A Review on Challenges and the State of the Art". In: *Renewable and Sustainable Energy Reviews* 160, p. 112246. ISSN: 1364-0321. DOI: 10.1016/j.rser.2022.112246. URL: <https://www.sciencedirect.com/science/article/pii/S136403212200168X>.
- Forbes Home (May 2024). *How Much Does The Tesla Powerwall Cost? (2024)*. URL: <https://web.archive.org/web/20240511112531/https://www.forbes.com/home-improvement/solar/tesla-powerwall-solar-battery-review/>.
- Furfari, Samuel and Alessandro Clerici (May 2021). "Green Hydrogen: The Crucial Performance of Electrolysers Fed by Variable and Intermittent Renewable Electricity". In: *The European Physical Journal Plus* 136.5, p. 509. ISSN: 2190-5444. DOI: 10.1140/epjp/s13360-021-01445-5. URL: <https://doi.org/10.1140/epjp/s13360-021-01445-5>.
- Fürsch, Michaela, Stephan Nagl, and Dietmar Lindenberger (Mar. 2014). "Optimization of Power Plant Investments under Uncertain Renewable Energy Deployment Paths: A Multistage Stochastic Programming Approach". In: *Energy Systems* 5.1, pp. 85–121. ISSN: 1868-3975. DOI: 10.1007/s12667-013-0094-0. URL: <https://doi.org/10.1007/s12667-013-0094-0>.
- Gabrielli, Paolo, Florian Fürer, Georgios Mavromatidis, and Marco Mazzotti (Mar. 2019). "Robust and Optimal Design of Multi-Energy Systems with Seasonal Storage through Uncertainty Analysis". In: *Applied Energy* 238, pp. 1192–1210. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.01.064. URL: <https://www.sciencedirect.com/science/article/pii/S0306261919300649>.
- Gagnon, Pieter, Robert Margolis, Jennifer Melius, Caleb Phillips, and Ryan Elmore (Feb. 2018). "Estimating Rooftop Solar Technical Potential across the US Using a Combination of GIS-based Methods, Lidar Data, and Statistical Modeling". In: *Environmental Research Letters* 13.2, p. 024027. ISSN: 1748-9326. DOI: 10.1088/1748-9326/aaa554. URL: <https://dx.doi.org/10.1088/1748-9326/aaa554>.
- Gang, Wenjie, Shengwei Wang, Kui Shan, and Diance Gao (May 2015). "Impacts of Cooling Load Calculation Uncertainties on the Design Optimization of Building Cooling Systems". In: *Energy and Buildings* 94, pp. 1–9. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.02.032. URL: <https://www.sciencedirect.com/science/article/pii/S0378778815001371>.
- Gawley, David and Paul McKenzie (June 2022). "Investigating the Suitability of GIS and Remotely-Sensed Datasets for Photovoltaic Modelling on Building Rooftops". In: *Energy and Buildings* 265, p. 112083. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2022.112083. URL: <https://www.sciencedirect.com/science/article/pii/S0378778822002547>.
- Gioia, Daniele (June 2023). *ScenarioReducer*. URL: <https://github.com/DanieleGioia/ScenarioReducer>.
- Girolami, Mark (Jan. 2020). "Introducing Data-Centric Engineering: An Open Access Journal Dedicated to the Transformation of Engineering Design and Practice". In: *Data-Centric Engineering* 1, e1. ISSN: 2632-6736. DOI: 10.1017/dce.2020.5. URL: <https://www.cambridge.org/core/journal>

- s/data-centric-engineering/article/introducing-datacentric-engineering-an-open-access-journal-dedicated-to-the-transformation-of-engineering-design-and-practice/4DDDC6D24B40C270364746CB2A5B2CF0.
- Gkantonas, Savvas, Daniel Zabotti, Leo CC Mesquita, Epaminondas Mastorakos, and Pedro M. de Oliveira (July 2021). "Airborne.Cam: A Risk Calculator of SARS-CoV-2 Aerosol Transmission under Well-Mixed Ventilation Conditions". In: URL: <https://www.repository.cam.ac.uk/handle/1810/324994>.
- Glasgo, Brock, Chris Hendrickson, and Inês Lima Azevedo (Oct. 2017). "Assessing the Value of Information in Residential Building Simulation: Comparing Simulated and Actual Building Loads at the Circuit Level". In: *Applied Energy* 203, pp. 348–363. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2017.05.164](https://doi.org/10.1016/j.apenergy.2017.05.164). URL: <https://www.sciencedirect.com/science/article/pii/S0306261917307237>.
- González-Vidal, Aurora, Fernando Jiménez, and Antonio F. Gómez-Skarmeta (Aug. 2019). "A Methodology for Energy Multivariate Time Series Forecasting in Smart Buildings Based on Feature Selection". In: *Energy and Buildings* 196, pp. 71–82. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2019.05.021](https://doi.org/10.1016/j.enbuild.2019.05.021). URL: <https://www.sciencedirect.com/science/article/pii/S0378778818338775>.
- Gould, John P (May 1974). "Risk, Stochastic Preference, and the Value of Information". In: *Journal of Economic Theory* 8.1, pp. 64–84. ISSN: 0022-0531. doi: [10.1016/0022-0531\(74\)90006-4](https://doi.org/10.1016/0022-0531(74)90006-4). URL: <https://www.sciencedirect.com/science/article/pii/0022053174900064>.
- Griffith, B., N. Long, P. Torcellini, et al. (Mar. 2008). *Methodology for Modeling Building Energy Performance across the Commercial Sector*. Tech. rep. NREL/TP-550-41956. National Renewable Energy Lab. (NREL), Golden, CO (United States). doi: [10.2172/926505](https://doi.org/10.2172/926505). URL: <https://www.osti.gov/biblio/926505>.
- Grussing, Michael N. (Aug. 2018). "Optimized Building Component Assessment Planning Using a Value of Information Model". In: *Journal of Performance of Constructed Facilities* 32.4, p. 04018054. ISSN: 1943-5509. doi: [10.1061/\(ASCE\)CF.1943-5509.0001198](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001198). URL: <https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0001198>.
- Haben, Stephen and Sam Young (Dec. 2023). *ADVice: AI for Decarbonisation Challenges*. Tech. rep. The Alan Turing Institute. URL: https://www.turing.ac.uk/sites/default/files/2023-12/advice_ai_for_decarbonisation_challenges.pdf.
- Han, KwangHoon and Jensen Zhang (May 2020). "Energy-Saving Building System Integration with a Smart and Low-Cost Sensing/Control Network for Sustainable and Healthy Living Environments: Demonstration Case Study". In: *Energy and Buildings* 214, p. 109861. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2020.109861](https://doi.org/10.1016/j.enbuild.2020.109861). URL: <https://www.sciencedirect.com/science/article/pii/S037878819329573>.
- Han, Mengjie, Zhenwu Wang, and Xingxing Zhang (Jan. 2021). "An Approach to Data Acquisition for Urban Building Energy Modeling Using a Gaussian Mixture Model and Expectation-Maximization Algorithm". In: *Buildings* 11.1, p. 30. ISSN: 2075-5309. doi: [10.3390/buildings11010030](https://doi.org/10.3390/buildings11010030). URL: <https://www.mdpi.com/2075-5309/11/1/30>.
- Hawkins, Douglas M. (Jan. 2004). "The Problem of Overfitting". In: *Journal of Chemical Information and Computer Sciences* 44.1, pp. 1–12. ISSN: 0095-2338. doi: [10.1021/ci0342472](https://doi.org/10.1021/ci0342472). URL: <https://doi.org/10.1021/ci0342472>.
- Heinen, Steve, Pierluigi Mancarella, Ciara O'Dwyer, and Mark O'Malley (July 2018). "Heat Electrification: The Latest Research in Europe". In: *IEEE Power and Energy Magazine* 16.4, pp. 69–78.

- ISSN: 1558-4216. DOI: [10.1109/MPE.2018.2822867](https://doi.org/10.1109/MPE.2018.2822867). URL: <https://ieeexplore.ieee.org/abstract/document/8386884>.
- Heitsch, Holger and Werner Römisch (Feb. 2003). "Scenario Reduction Algorithms in Stochastic Programming". In: *Computational Optimization and Applications* 24.2, pp. 187–206. ISSN: 1573-2894. DOI: [10.1023/A:1021805924152](https://doi.org/10.1023/A:1021805924152). URL: <https://doi.org/10.1023/A:1021805924152>.
- Hernández, José L., Ignacio de Miguel, Fredy Vélez, and Ali Vasallo (July 2024). "Challenges and Opportunities in European Smart Buildings Energy Management: A Critical Review". In: *Renewable and Sustainable Energy Reviews* 199, p. 114472. ISSN: 1364-0321. DOI: [10.1016/j.ser.2024.114472](https://doi.org/10.1016/j.ser.2024.114472). URL: <https://www.sciencedirect.com/science/article/pii/S1364032124001953>.
- Hörsch, Jonas, Fabian Hofmann, David Schlachtberger, and Tom Brown (Nov. 2018). "PyPSA-Eur: An Open Optimisation Model of the European Transmission System". In: *Energy Strategy Reviews* 22, pp. 207–215. ISSN: 2211-467X. DOI: [10.1016/j.esr.2018.08.012](https://doi.org/10.1016/j.esr.2018.08.012). URL: <https://www.sciencedirect.com/science/article/pii/S2211467X18300804>.
- House of Commons (Feb. 2022). *Decarbonising Heat in Homes: Seventh Report of Session 2021–22*. Tech. rep. House of Commons Business, Energy and Industrial Strategy Committee. URL: <https://committees.parliament.uk/publications/8742/documents/88647/default/>.
- Howard, Ronald A. (Aug. 1966). "Information Value Theory". In: *IEEE Transactions on Systems Science and Cybernetics* 2.1, pp. 22–26. ISSN: 2168-2887. DOI: [10.1109/TSSC.1966.300074](https://doi.org/10.1109/TSSC.1966.300074). URL: <https://ieeexplore.ieee.org/abstract/document/4082064>.
- Hu, Maomao and Fu Xiao (Mar. 2020). "Quantifying Uncertainty in the Aggregate Energy Flexibility of High-Rise Residential Building Clusters Considering Stochastic Occupancy and Occupant Behavior". In: *Energy* 194, p. 116838. ISSN: 0360-5442. DOI: [10.1016/j.energy.2019.116838](https://doi.org/10.1016/j.energy.2019.116838). URL: <https://www.sciencedirect.com/science/article/pii/S0360544219325332>.
- Hu, Yuanwei, Bo Yang, Pengyu Wu, et al. (Oct. 2024). "Optimal Planning of Electric-Heating Integrated Energy System in Low-Carbon Park with Energy Storage System". In: *Journal of Energy Storage* 99, p. 113327. ISSN: 2352-152X. DOI: [10.1016/j.est.2024.113327](https://doi.org/10.1016/j.est.2024.113327). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X2402913X>.
- Hurley, Wesley V. (Jan. 1964). "A Mathematical Theory of the Value of Information". In: *International Journal of Computer Mathematics* 1.1-4, pp. 97–146. ISSN: 0020-7160, 1029-0265. DOI: [10.1080/00207167508803009](https://doi.org/10.1080/00207167508803009). URL: <http://www.tandfonline.com/doi/abs/10.1080/00207167508803009>.
- IEA (Oct. 2021). *Net Zero by 2050 - A Roadmap for the Global Energy Sector*. Tech. rep. International Energy Agency. URL: https://iea.blob.core.windows.net/assets/deebef5d-0c34-4539-9d0c-10b13d840027/NetZeroBy2050-ARoadmapfortheGlobalEnergySector_CORR.pdf.
- (July 2023). *Tracking Clean Energy Progress 2023*. Tech. rep. IEA, Paris. URL: <https://www.iea.org/reports/tracking-clean-energy-progress-2023>.
- Inc, Hydrostor (Sept. 2023). *Hydrostor – Written Evidence (LES0037)*. Tech. rep. UK Government Science and Technology Committee. URL: <https://committees.parliament.uk/writtenevidence/124537/pdf/>.
- IRENA (Oct. 2017). *Electricity Storage and Renewables: Costs and Markets to 2030*. Tech. rep. International Renewable Energy Agency. URL: <https://www.irena.org/publications/2017/Oct/Electricity-storage-and-renewables-costs-and-markets>.

- IRENA (Aug. 2023). *Renewable Power Generation Costs in 2022*. Tech. rep. International Renewable Energy Agency. URL: <https://www.irena.org/Publications/2023/Aug/Renewable-Power-Generation-Costs-in-2022>.
- Iweh, Chu Donatus, Samuel Gyamfi, Emmanuel Tanyi, and Eric Effah-Donyina (Jan. 2021). “Distributed Generation and Renewable Energy Integration into the Grid: Prerequisites, Push Factors, Practical Options, Issues and Merits”. In: *Energies* 14.17, p. 5375. ISSN: 1996-1073. doi: 10.3390/en14175375. URL: <https://www.mdpi.com/1996-1073/14/17/5375>.
- Jane Fennell, Pamela, Paul A Ruyssevelt, Érika Mata, and Martin Jakob (2019). “A Review of the Status of Uncertainty and Sensitivity Analysis in Urban Building Energy Models”. In: *Building Simulation 2019*. Vol. 16. Building Simulation. IBPSA, pp. 3353–3360. ISBN: 978-1-77505-201-2. doi: 10.26868/25222708.2019.210478. URL: https://publications.ibpsa.org/conference/paper/?id=bs2019_210478.
- Ji, Zhengxiong, Jianyan Tian, Shuwei Liu, et al. (Jan. 2025). “Optimal Scheduling of Park-Level Integrated Energy System Considering Multiple Uncertainties: A Comprehensive Risk Strategy-Information Gap Decision Theory Method”. In: *Applied Energy* 377, p. 124700. ISSN: 0306-2619. doi: 10.1016/j.apenergy.2024.124700. URL: <https://www.sciencedirect.com/science/article/pii/S030626192402083X>.
- Kang, Xuyuan, Jingjing An, and Da Yan (Feb. 2023). “A Systematic Review of Building Electricity Use Profile Models”. In: *Energy and Buildings* 281, p. 112753. ISSN: 0378-7788. doi: 10.1016/j.enbuild.2022.112753. URL: <https://www.sciencedirect.com/science/article/pii/S0378778822009240>.
- Kathirgamanathan, Anjukan, Mattia De Rosa, Eleni Mangina, and Donal P. Finn (Jan. 2021). “Data-Driven Predictive Control for Unlocking Building Energy Flexibility: A Review”. In: *Renewable and Sustainable Energy Reviews* 135, p. 110120. ISSN: 1364-0321. doi: 10.1016/j.rser.2020.110120. URL: <https://www.sciencedirect.com/science/article/pii/S1364032120304111>.
- Kebede, Abraham Alem, Theodoros Kalogiannis, Joeri Van Mierlo, and Maitane Berecibar (May 2022). “A Comprehensive Review of Stationary Energy Storage Devices for Large Scale Renewable Energy Sources Grid Integration”. In: *Renewable and Sustainable Energy Reviews* 159, p. 112213. ISSN: 1364-0321. doi: 10.1016/j.rser.2022.112213. URL: <https://www.sciencedirect.com/science/article/pii/S1364032122001368>.
- Keisler, Jeffrey M., Zachary A. Collier, Eric Chu, Nina Sinatra, and Igor Linkov (Mar. 2014). “Value of Information Analysis: The State of Application”. In: *Environment Systems and Decisions* 34.1, pp. 3–23. ISSN: 2194-5411. doi: 10.1007/s10669-013-9439-4. URL: <https://doi.org/10.1007/s10669-013-9439-4>.
- Kensa (Mar. 2014). *How COP Varies with Inlet and Outlet Temperature V2*. Tech. rep. URL: <https://www.kensaheatpumps.com/wp-content/uploads/2014/03/Factsheet-COP-Variation-V2.pdf>.
- Keyvandarian, Ali and Ahmed Saif (Mar. 2023). “Optimal Sizing of a Reliability-Constrained, Stand-Alone Hybrid Renewable Energy System Using Robust Satisficing”. In: *Renewable Energy* 204, pp. 569–579. ISSN: 0960-1481. doi: 10.1016/j.renene.2023.01.036. URL: <https://www.sciencedirect.com/science/article/pii/S0960148123000447>.
- Khezri, Rahmat, Amin Mahmoudi, and Hirohisa Aki (Jan. 2022). “Optimal Planning of Solar Photovoltaic and Battery Storage Systems for Grid-Connected Residential Sector: Review, Challenges and New Perspectives”. In: *Renewable and Sustainable Energy Reviews* 153, p. 111763. ISSN: 1364-0321. doi: 10.1016/j.rser.2021.111763. URL: <https://www.sciencedirect.com/science/article/pii/S1364032121010339>.

- Kim, Jangkyum, Yohwan Choi, Seunghyoung Ryu, and Hongseok Kim (Apr. 2017). "Robust Operation of Energy Storage System with Uncertain Load Profiles". In: *Energies* 10.4, p. 416. ISSN: 1996-1073. doi: [10.3390/en10040416](https://doi.org/10.3390/en10040416). URL: <https://www.mdpi.com/1996-1073/10/4/416>.
- Kim, Sunwoo, Yechan Choi, Joungho Park, et al. (Feb. 2024). "Multi-Period, Multi-Timescale Stochastic Optimization Model for Simultaneous Capacity Investment and Energy Management Decisions for Hybrid Micro-Grids with Green Hydrogen Production under Uncertainty". In: *Renewable and Sustainable Energy Reviews* 190, p. 114049. ISSN: 1364-0321. doi: [10.1016/j.rser.2023.114049](https://doi.org/10.1016/j.rser.2023.114049). URL: <https://www.sciencedirect.com/science/article/pii/S1364032123009073>.
- Kim, Woohyun, Yerim Han, Kyoung Jae Kim, and Kwan-Woo Song (Nov. 2020). "Electricity Load Forecasting Using Advanced Feature Selection and Optimal Deep Learning Model for the Variable Refrigerant Flow Systems". In: *Energy Reports* 6, pp. 2604–2618. ISSN: 2352-4847. doi: [10.1016/j.egyr.2020.09.019](https://doi.org/10.1016/j.egyr.2020.09.019). URL: <https://www.sciencedirect.com/science/article/pii/S2352484720313317>.
- King, W., D. Banks, and J. Findlay (Nov. 2012). "Field Determination of Shallow Soil Thermal Conductivity Using a Short-Duration Needle Probe Test". In: *Quarterly Journal of Engineering Geology and Hydrogeology* 45.4, pp. 497–504. ISSN: 1470-9236. doi: [10.1144/qjegh2012-002](https://doi.org/10.1144/qjegh2012-002). URL: <https://doi.org/10.1144/qjegh2012-002>.
- Kingma, Diederik P. and Jimmy Ba (Jan. 2017). *Adam: A Method for Stochastic Optimization*. doi: [10.48550/arXiv.1412.6980](https://arxiv.org/abs/1412.6980). arXiv: [1412.6980 \[cs\]](https://arxiv.org/abs/1412.6980). URL: <http://arxiv.org/abs/1412.6980>.
- Kittner, Noah, Oliver Schmidt, Iain Staffell, and Daniel M. Kammen (Jan. 2020). "Chapter 8 - Grid-scale Energy Storage". In: *Technological Learning in the Transition to a Low-Carbon Energy System*. Ed. by Martin Junginger and Atse Louwen. Academic Press, pp. 119–143. ISBN: 978-0-12-818762-3. doi: [10.1016/B978-0-12-818762-3.00008-X](https://doi.org/10.1016/B978-0-12-818762-3.00008-X). URL: <https://www.sciencedirect.com/science/article/pii/B978012818762300008X>.
- Klemeš, V. (1977). "Value of Information in Reservoir Optimization". In: *Water Resources Research* 13.5, pp. 837–850. ISSN: 1944-7973. doi: [10.1029/WR013i005p00837](https://doi.org/10.1029/WR013i005p00837). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/WR013i005p00837>.
- Krukanont, Pongsak and Tetsuo Tezuka (Oct. 2007). "Implications of Capacity Expansion under Uncertainty and Value of Information: The near-Term Energy Planning of Japan". In: *Energy* 32.10, pp. 1809–1824. ISSN: 0360-5442. doi: [10.1016/j.energy.2007.02.003](https://doi.org/10.1016/j.energy.2007.02.003). URL: <https://www.sciencedirect.com/science/article/pii/S0360544207000333>.
- Laboratory, Basalt Ridge Testing (Mar. 2024). *Basalt Ridge Testing Laboratory Price List*. Tech. rep. URL: <https://web.archive.org/web/20240305171909/https://www.basaltridgetesting.com/pricing>.
- Lamarche, Louis and Benoit Beauchamp (Feb. 2007). "A New Contribution to the Finite Line-Source Model for Geothermal Boreholes". In: *Energy and Buildings* 39.2, pp. 188–198. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2006.06.003](https://doi.org/10.1016/j.enbuild.2006.06.003). URL: <https://www.sciencedirect.com/science/article/pii/S0378778806001824>.
- Langtry, Max** and Ruchi Choudhary (Apr. 2024). *Cambridge University Estates Building Energy Usage Archive*. doi: [10.5281/zenodo.10955332](https://doi.org/10.5281/zenodo.10955332). URL: <https://github.com/EECi/Cambridge-Estates-Building-Energy-Archive>.
- (Dec. 2025a). "Quantifying the Benefit of Load Uncertainty Reduction for the Design of District Energy Systems under Grid Constraints Using the Value of Information". In: *Applied Energy* 400, p. 126549. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2025.126549](https://doi.org/10.1016/j.apenergy.2025.126549). URL: <https://www.sciencedirect.com/science/article/pii/S0306261925012796>.

- Langtry, Max** and Ruchi Choudhary (Oct. 2025b). "The Value of Hedging against Energy Storage Uncertainties When Designing Energy Parks". In: *Energy* 334, p. 137600. ISSN: 0360-5442. doi: [10.1016/j.energy.2025.137600](https://doi.org/10.1016/j.energy.2025.137600). URL: <https://www.sciencedirect.com/science/article/pii/S0360544225032426>.
- Langtry, Max**, Vija Wichitwechkarn, Rebecca Ward, et al. (Oct. 2024). "Impact of Data for Forecasting on Performance of Model Predictive Control in Buildings with Smart Energy Storage". In: *Energy and Buildings* 320, p. 114605. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2024.114605](https://doi.org/10.1016/j.enbuild.2024.114605). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824007217>.
- Langtry, Max**, Chaoqun Zhuang, Rebecca Ward, et al. (Nov. 2024). "Rationalising Data Collection for Supporting Decision Making in Building Energy Systems Using Value of Information Analysis". In: *Journal of Building Performance Simulation*, pp. 1–17. ISSN: 1940-1493. doi: [10.1080/19401493.2024.2423827](https://doi.org/10.1080/19401493.2024.2423827). URL: <https://doi.org/10.1080/19401493.2024.2423827>.
- (2023). "Value of Information Analysis for Rationalising Information Gathering in Building Energy Analysis". In: *Building Simulation 2023*. Vol. 18. Building Simulation. IBPSA, pp. 3483–3490. doi: [10.26868/25222708.2023.1478](https://doi.org/10.26868/25222708.2023.1478). URL: https://publications.ibpsa.org/conference/paper/?id=bs2023_1478.
- Lee, Doyun, Ryozo Ooka, Shintaro Ikeda, Wonjun Choi, and Younghoon Kwak (Oct. 2020). "Model Predictive Control of Building Energy Systems with Thermal Energy Storage in Response to Occupancy Variations and Time-Variant Electricity Prices". In: *Energy and Buildings* 225, p. 110291. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2020.110291](https://doi.org/10.1016/j.enbuild.2020.110291). URL: <https://www.sciencedirect.com/science/article/pii/S0378778820302875>.
- Leibowicz, Benjamin D., Christopher M. Lanham, Max T. Brozynski, et al. (Nov. 2018). "Optimal Decarbonization Pathways for Urban Residential Building Energy Services". In: *Applied Energy* 230, pp. 1311–1325. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2018.09.046](https://doi.org/10.1016/j.apenergy.2018.09.046). URL: <https://www.sciencedirect.com/science/article/pii/S0306261918313552>.
- Li, Benjia, Zhongbing Liu, Yaling Wu, et al. (May 2023). "Review on Photovoltaic with Battery Energy Storage System for Power Supply to Buildings: Challenges and Opportunities". In: *Journal of Energy Storage* 61, p. 106763. ISSN: 2352-152X. doi: [10.1016/j.est.2023.106763](https://doi.org/10.1016/j.est.2023.106763). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X23001603>.
- Li, Peifeng, Wei Wang, Jun Wei, et al. (Nov. 2021). "Optimal Operation Model of Park Integrated Energy System Considering Uncertainty of Flexible Electrical, Thermal and Gas Load". In: *Journal of Physics: Conference Series* 2095.1, p. 012025. ISSN: 1742-6596. doi: [10.1088/1742-6596/2095/1/012025](https://doi.org/10.1088/1742-6596/2095/1/012025). URL: <https://dx.doi.org/10.1088/1742-6596/2095/1/012025>.
- Li, X., H. Arbabi, G. Bennett, T. Oreszczyn, and D. Densley Tingley (June 2022). "Net Zero by 2050: Investigating Carbon-Budget Compliant Retrofit Measures for the English Housing Stock". In: *Renewable and Sustainable Energy Reviews* 161, p. 112384. ISSN: 1364-0321. doi: [10.1016/j.rser.2022.112384](https://doi.org/10.1016/j.rser.2022.112384). URL: <https://www.sciencedirect.com/science/article/pii/S1364032122002945>.
- Li, Xiwang and Jin Wen (Sept. 2014). "Review of Building Energy Modeling for Control and Operation". In: *Renewable and Sustainable Energy Reviews* 37, pp. 517–537. ISSN: 1364-0321. doi: [10.1016/j.rser.2014.05.056](https://doi.org/10.1016/j.rser.2014.05.056). URL: <https://www.sciencedirect.com/science/article/pii/S1364032114003815>.
- Lim, Bryan, Sercan Ö. Arik, Nicolas Loeff, and Tomas Pfister (Oct. 2021). "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting". In: *International Journal of Forecasting* 37.4, pp. 1748–1764. ISSN: 0169-2070. doi: [10.1016/j.ijforecast.2021.03.012](https://doi.org/10.1016/j.ijforecast.2021.03.012). URL: <https://www.sciencedirect.com/science/article/pii/S0169207021000637>.

- Liu, Bo and Yufei Wang (Jan. 2020). "6 - Energy System Optimization under Uncertainties: A Comprehensive Review". In: *Towards Sustainable Chemical Processes*. Ed. by Jingzheng Ren, Yufei Wang, and Chang He. Elsevier, pp. 149–170. ISBN: 978-0-12-818376-2. doi: [10.1016/B978-0-12-818376-2.00006-5](https://doi.org/10.1016/B978-0-12-818376-2.00006-5). URL: <https://www.sciencedirect.com/science/article/pii/B9780128183762000065>.
- López Santos, Miguel, Xela García-Santiago, Fernando Echevarría Camarero, Gonzalo Blázquez Gil, and Pablo Carrasco Ortega (Jan. 2022). "Application of Temporal Fusion Transformer for Day-Ahead PV Power Forecasting". In: *Energies* 15.14, p. 5232. ISSN: 1996-1073. doi: [10.3390/en15145232](https://doi.org/10.3390/en15145232). URL: <https://www.mdpi.com/1996-1073/15/14/5232>.
- Loveridge, Fleur, Gary Holmes, William Powrie, and Toby Roberts (Apr. 2013). "Thermal Response Testing through the Chalk Aquifer in London, UK". In: *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering* 166.2, pp. 197–210. ISSN: 1353-2618, 1751-8563. doi: [10.1680/geng.12.00037](https://doi.org/10.1680/geng.12.00037). URL: <https://www.icevirtuallibrary.com/doi/10.1680/geng.12.00037>.
- Low, J E, Fleur Loveridge, and William Powrie (2013). "Measuring Soil Thermal Properties for Use in Energy Foundation Design". In: *Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris 2013*, pp. 3375–3378. URL: <https://www.cfms-sols.org/sites/default/files/Actes/3375-3378.pdf>.
- Lu, Jie, Chaobo Zhang, Bozheng Li, et al. (Feb. 2025). "Self-Attention Variational Autoencoder-Based Method for Incomplete Model Parameter Imputation of Digital Twin Building Energy Systems". In: *Energy and Buildings* 328, p. 115162. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2024.115162](https://doi.org/10.1016/j.enbuild.2024.115162). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824012787>.
- Lu, Jie, Zeyu Zheng, **Max Langtry**, et al. (July 2025). *Automated Building Energy Modeling for Energy Retrofits Using a Large Language Model-Based Multi-Agent Framework*. (Submitted to iScience).
- Luo, Na, Zhe Wang, David Blum, et al. (Apr. 2022). "A Three-Year Dataset Supporting Research on Building Energy Management and Occupancy Analytics". In: *Scientific Data* 9.1, p. 156. ISSN: 2052-4463. doi: [10.1038/s41597-022-01257-x](https://doi.org/10.1038/s41597-022-01257-x). URL: <https://www.nature.com/articles/s41597-022-01257-x>.
- Macdonald, Iain and Paul Strachan (Feb. 2001). "Practical Application of Uncertainty Analysis". In: *Energy and Buildings* 33.3, pp. 219–227. ISSN: 0378-7788. doi: [10.1016/S0378-7788\(00\)00085-2](https://doi.org/10.1016/S0378-7788(00)00085-2). URL: <https://www.sciencedirect.com/science/article/pii/S0378778800000852>.
- MacKay, David JC (May 2016). *Sustainable Energy - without the Hot Air*. Bloomsbury Publishing. ISBN: 978-1-906860-45-5.
- Maia, Leonardo K. K., Anton Ochoa Bique, and Edwin Zondervan (Jan. 2021). "Certainty through Uncertainty: Stochastic Optimization of Grid-Integrated Large-Scale Energy Storage in Germany". In: *Physical Sciences Reviews* 8.1, pp. 103–120. ISSN: 2365-659X. doi: [10.1515/psr-2020-0051](https://doi.org/10.1515/psr-2020-0051). URL: <https://www.degruyter.com/document/doi/10.1515/psr-2020-0051/html>.
- Majidi, M., B. Mohammadi-Ivatloo, and A. Soroudi (Sept. 2019). "Application of Information Gap Decision Theory in Practical Energy Problems: A Comprehensive Review". In: *Applied Energy* 249, pp. 157–165. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2019.04.144](https://doi.org/10.1016/j.apenergy.2019.04.144). URL: <https://www.sciencedirect.com/science/article/pii/S0306261919308037>.
- Malings, Carl and Matteo Pozzi (Oct. 2016). "Value of Information for Spatially Distributed Systems: Application to Sensor Placement". In: *Reliability Engineering & System Safety* 154, pp. 219–233.

- ISSN: 0951-8320. doi: [10.1016/j.ress.2016.05.010](https://doi.org/10.1016/j.ress.2016.05.010). URL: <https://www.sciencedirect.com/science/article/pii/S0951832016300564>.
- Mantha, Bharadwaj, Chen Feng, Carol Menassa, and Vineet Kamat (June 2015). "Real-Time Building Energy and Comfort Parameter Data Collection Using Mobile Indoor Robots". In: *ISARC Proceedings*, pp. 1–9. URL: https://www.iaarc.org/publications/2015_proceedings_of_the_32st_isarc_oulu_finland/real_time_building_energy_and_comfort_parameter_data_collection_using_mobile_indoor_robots.html.
- Mavromatidis, Georgios, Kristina Orehounig, and Jan Carmeliet (May 2018a). "A Review of Uncertainty Characterisation Approaches for the Optimal Design of Distributed Energy Systems". In: *Renewable and Sustainable Energy Reviews* 88, pp. 258–277. ISSN: 1364-0321. doi: [10.1016/j.rser.2018.02.021](https://doi.org/10.1016/j.rser.2018.02.021). URL: <https://www.sciencedirect.com/science/article/pii/S1364032118300510>.
- (Mar. 2018b). "Uncertainty and Global Sensitivity Analysis for the Optimal Design of Distributed Energy Systems". In: *Applied Energy* 214, pp. 219–238. ISSN: 0306-2619. doi: [10.1016/j.apene.2018.01.062](https://doi.org/10.1016/j.apene.2018.01.062). URL: <https://www.sciencedirect.com/science/article/pii/S0306261918300710>.
- Mazzoni, Stefano, Sean Ooi, Benedetto Nastasi, and Alessandro Romagnoli (Nov. 2019). "Energy Storage Technologies as Techno-Economic Parameters for Master-Planning and Optimal Dispatch in Smart Multi Energy Systems". In: *Applied Energy* 254, p. 113682. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2019.113682](https://doi.org/10.1016/j.apenergy.2019.113682). URL: <https://www.sciencedirect.com/science/article/pii/S0306261919313698>.
- Met Office (2022). *MIDAS Open: UK Hourly Weather Observation Data, V202207*. doi: [10.5285/6180FB7ED76A442EB1B8F3F152FD08D7](https://doi.org/10.5285/6180FB7ED76A442EB1B8F3F152FD08D7). URL: <https://catalogue.ceda.ac.uk/uuid/6180fb7ed76a442eb1b8f3f152fd08d7>.
- Min, Daiki, Jong-hyun Ryu, and Dong Gu Choi (Aug. 2018). "A Long-Term Capacity Expansion Planning Model for an Electric Power System Integrating Large-Size Renewable Energy Technologies". In: *Computers & Operations Research* 96, pp. 244–255. ISSN: 0305-0548. doi: [10.1016/j.cor.2017.10.006](https://doi.org/10.1016/j.cor.2017.10.006). URL: <https://www.sciencedirect.com/science/article/pii/S0305054817302642>.
- Mitchell, Rachel and Sukumar Natarajan (Oct. 2020). "UK Passivhaus and the Energy Performance Gap". In: *Energy and Buildings* 224, p. 110240. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2020.110240](https://doi.org/10.1016/j.enbuild.2020.110240). URL: <https://www.sciencedirect.com/science/article/pii/S0378778820313918>.
- Mobaraki, Behnam, Seyedmilad Komarizadehasl, Francisco Javier Castilla Pascual, José Antonio Lozano-Galant, and Rocio Porras Soriano (May 2022). "A Novel Data Acquisition System for Obtaining Thermal Parameters of Building Envelopes". In: *Buildings* 12.5, p. 670. ISSN: 2075-5309. doi: [10.3390/buildings12050670](https://doi.org/10.3390/buildings12050670). URL: <https://www.mdpi.com/2075-5309/12/5/670>.
- Mock, Theodore J. (1971). "Concepts of Information Value and Accounting". In: *The Accounting Review* 46.4, pp. 765–778. ISSN: 0001-4826. JSTOR: [244255](https://doi.org/10.2307/244255). URL: <https://www.jstor.org/stable/244255>.
- Modiano, Eduardo Marco (1987). "Derived Demand and Capacity Planning under Uncertainty". In: *Operations Research* 35.2, pp. 185–197. ISSN: 0030-364X. JSTOR: [170690](https://doi.org/10.1287/opre.35.2.185). URL: <https://www.jstor.org/stable/170690>.
- Mohammadi, Fatemeh, Faramarz Faghihi, Ahad Kazemi, and Amir H. Salemi (Aug. 2022). "The Effect of Multi -Uncertainties on Battery Energy Storage System Sizing in Smart Homes". In: *Journal of Energy Storage* 52, p. 104765. ISSN: 2352-152X. doi: [10.1016/j.est.2022.104765](https://doi.org/10.1016/j.est.2022.104765). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X22007757>.

- Mohseni, Soheil and Alan C. Brent (July 2022). "Quantifying the Effects of Forecast Uncertainty on the Role of Different Battery Technologies in Grid-Connected Solar Photovoltaic/Wind/Micro-Hydro Micro-Grids: An Optimal Planning Study". In: *Journal of Energy Storage* 51, p. 104412. ISSN: 2352-152X. doi: [10.1016/j.est.2022.104412](https://doi.org/10.1016/j.est.2022.104412). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X22004364>.
- Molina-Solana, Miguel, María Ros, M. Dolores Ruiz, Juan Gómez-Romero, and M. J. Martin-Bautista (Apr. 2017). "Data Science for Building Energy Management: A Review". In: *Renewable and Sustainable Energy Reviews* 70, pp. 598–609. ISSN: 1364-0321. doi: [10.1016/j.rser.2016.11.132](https://doi.org/10.1016/j.rser.2016.11.132). URL: <https://www.sciencedirect.com/science/article/pii/S1364032116308814>.
- Motegi, Naoya, Mary Ann Piette, Satkartar Kinney, and Jim Dewey (Sept. 2003). "Case Studies of Energy Information Systems and Related Technology: Operational Practices, Costs, and Benefits". In: *Lawrence Berkeley National Laboratory*. URL: <https://escholarship.org/content/qt49t8d6dc/qt49t8d6dc.pdf>.
- Mott MacDonald (Aug. 2018). *Storage Cost and Technical Assumptions for BEIS: Summary Document*. Tech. rep. URL: <https://assets.publishing.service.gov.uk/media/5f3cf6c9d3bf7f1b0fa7a165/storage-costs-technical-assumptions-2018.pdf>.
- Mu, Yunfei, Congshan Wang, Yan Cao, et al. (May 2022). "A CVaR-based Risk Assessment Method for Park-Level Integrated Energy System Considering the Uncertainties and Correlation of Energy Prices". In: *Energy* 247, p. 123549. ISSN: 0360-5442. doi: [10.1016/j.energy.2022.123549](https://doi.org/10.1016/j.energy.2022.123549). URL: <https://www.sciencedirect.com/science/article/pii/S0360544222004522>.
- Myklebust, Hans Olav Vogt, Jo Eidsvik, Iver Bakken Sperstad, and Debarun Bhattacharjya (June 2020). "Value of Information Analysis for Complex Simulator Models: Application to Wind Farm Maintenance". In: *Decision Analysis* 17.2, pp. 134–153. ISSN: 1545-8490. doi: [10.1287/deca.2019.0405](https://doi.org/10.1287/deca.2019.0405). URL: <https://pubsonline.informs.org/doi/abs/10.1287/deca.2019.0405>.
- Nagy, Zoltan, Kingsley Nweye, Sharada Mohanty, et al. (Dec. 2023). *The CityLearn Challenge 2023, NeurIPS*. URL: <https://neurips.cc/virtual/2023/competition/66590>.
- National Grid ESO (Mar. 2020). *Historic Generation Mix & Carbon Intensity*. URL: <https://data.nationgrideso.com/carbon-intensity1/historic-generation-mix>.
- NESO (July 2022). *Potential Electricity Storage Routes to 2050*. Tech. rep. National Grid Electricity System Operator. URL: <https://www.neso.energy/document/273166/download>.
- Networks, Eastern Power (Nov. 2023). *Eastern Power Networks Plc: Use of System Charging Statement*. Tech. rep. UK Power Networks. URL: <https://media.umbraco.io/uk-power-networks/gwcflqcq/eastern-power-networks-lc14-statement-2023-v1-4.pdf>.
- Neumann, Fabian and Tom Brown (Jan. 2021). "The Near-Optimal Feasible Space of a Renewable Power System Model". In: *Electric Power Systems Research* 190, p. 106690. ISSN: 0378-7796. doi: [10.1016/j.epsr.2020.106690](https://doi.org/10.1016/j.epsr.2020.106690). URL: <https://www.sciencedirect.com/science/article/pii/S0378779620304934>.
- Niu, Jide, Xiaoyuan Li, Zhe Tian, and Hongxing Yang (Nov. 2023). "A Framework for Quantifying the Value of Information to Mitigate Risk in the Optimal Design of Distributed Energy Systems under Uncertainty". In: *Applied Energy* 350, p. 121717. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2023.121717](https://doi.org/10.1016/j.apenergy.2023.121717). URL: <https://www.sciencedirect.com/science/article/pii/S0306261923010814>.
- Niveditha, N. and M. M. Rajan Singaravel (Oct. 2022). "Optimal Sizing of Hybrid PV–Wind–Battery Storage System for Net Zero Energy Buildings to Reduce Grid Burden". In: *Applied Energy* 324,

- p. 119713. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2022.119713](https://doi.org/10.1016/j.apenergy.2022.119713). URL: <https://www.sciencedirect.com/science/article/pii/S0306261922010054>.
- Nouvel, Romain, Mariela Cotrado, and Dirk Pietruschka (2015). "European Mapping of Seasonal Performances of Air-Source and Geothermal Heat Pumps for Residential Applications". In: *Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale*.
- Novoa, Laura, Robert Flores, and Jack Brouwer (Dec. 2019). "Optimal Renewable Generation and Battery Storage Sizing and Siting Considering Local Transformer Limits". In: *Applied Energy* 256, p. 113926. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2019.113926](https://doi.org/10.1016/j.apenergy.2019.113926). URL: <https://www.sciencedirect.com/science/article/pii/S0306261919316137>.
- Nweye, Kingsley, Zoltan Nagy, Sharada Mohanty, et al. (Aug. 2023). "The CityLearn Challenge 2022: Overview, Results, and Lessons Learned". In: *Proceedings of the NeurIPS 2022 Competitions Track*. PMLR, pp. 85–103. URL: <https://proceedings.mlr.press/v220/nweye23a.html>.
- O'Neill, Zheng and Fuxin Niu (Apr. 2017). "Uncertainty and Sensitivity Analysis of Spatio-Temporal Occupant Behaviors on Residential Building Energy Usage Utilizing Karhunen-Loèvre Expansion". In: *Building and Environment* 115, pp. 157–172. ISSN: 0360-1323. doi: [10.1016/j.buildenv.2017.01.025](https://doi.org/10.1016/j.buildenv.2017.01.025). URL: <https://www.sciencedirect.com/science/article/pii/S0360132317300380>.
- O'Shaughnessy, Eric and Monisha Shah (Sept. 2021). *The Demand-Side Opportunity: The Roles of Distributed Solar and Building Energy Systems in a Decarbonized Grid*. Tech. rep. NREL/TP-6A20-80527. National Renewable Energy Lab. (NREL), Golden, CO (United States); Clean Kilowatts, LLC, Plymouth, MI (United States). doi: [10.2172/1820102](https://doi.org/10.2172/1820102). URL: <https://www.osti.gov/biblio/1820102>.
- Oldewurtel, Frauke, Alessandra Parisio, Colin N. Jones, et al. (Feb. 2012). "Use of Model Predictive Control and Weather Forecasts for Energy Efficient Building Climate Control". In: *Energy and Buildings* 45, pp. 15–27. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2011.09.022](https://doi.org/10.1016/j.enbuild.2011.09.022). URL: <https://www.sciencedirect.com/science/article/pii/S0378778811004105>.
- Oldham, Max (July 1996). "A Value for Information". In: *CIA Historical Review Program*. URL: <https://www.cia.gov/resources/csi/static/Value-for-Information.pdf>.
- ONS (Mar. 2023). *Coronavirus (COVID-19) Latest Insights - Office for National Statistics*. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsandiseases/articles/coronaviruscovid19latestinsights/infections>.
- (Oct. 2022). *Employee Earnings in the UK - Office for National Statistics*. URL: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2022>.
- Open Energy Modelling Initiative Wiki* (Jan. 2024). URL: https://wiki.openmod-initiative.org/wiki/Main_Page.
- Pang, Zhihong, Zheng O'Neill, Yanfei Li, and Fuxin Niu (Feb. 2020). "The Role of Sensitivity Analysis in the Building Performance Analysis: A Critical Review". In: *Energy and Buildings* 209, p. 109659. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2019.109659](https://doi.org/10.1016/j.enbuild.2019.109659). URL: <https://www.sciencedirect.com/science/article/pii/S0378778819319164>.
- Papadis, Elisa and George Tsatsaronis (Aug. 2020). "Challenges in the Decarbonization of the Energy Sector". In: *Energy* 205, p. 118025. ISSN: 0360-5442. doi: [10.1016/j.energy.2020.118025](https://doi.org/10.1016/j.energy.2020.118025). URL: <https://www.sciencedirect.com/science/article/pii/S0360544220311324>.

- Pedersen, Tim T., Marta Victoria, Morten G. Rasmussen, and Gorm B. Andresen (Nov. 2021). "Modeling All Alternative Solutions for Highly Renewable Energy Systems". In: *Energy* 234, p. 121294. ISSN: 0360-5442. doi: [10.1016/j.energy.2021.121294](https://doi.org/10.1016/j.energy.2021.121294). URL: <https://www.sciencedirect.com/science/article/pii/S0360544221015425>.
- Pelz, Peter F., Marc E. Pfetsch, Sebastian Kersting, et al. (2021). "Types of Uncertainty". In: *Mastering Uncertainty in Mechanical Engineering*. Ed. by Peter F. Pelz, Peter Groche, Marc E. Pfetsch, and Maximilian Schaeffner. Cham: Springer International Publishing, pp. 25–42. ISBN: 978-3-030-78354-9. doi: [10.1007/978-3-030-78354-9_2](https://doi.org/10.1007/978-3-030-78354-9_2). URL: https://doi.org/10.1007/978-3-030-78354-9_2.
- Perrin, Richard K. (1976). "The Value of Information and the Value of Theoretical Models in Crop Response Research". In: *American Journal of Agricultural Economics* 58.1, pp. 54–61. ISSN: 1467-8276. doi: [10.2307/1238577](https://doi.org/10.2307/1238577). URL: <https://onlinelibrary.wiley.com/doi/abs/10.2307/1238577>.
- Petkov, Ivalin and Paolo Gabrielli (Sept. 2020). "Power-to-Hydrogen as Seasonal Energy Storage: An Uncertainty Analysis for Optimal Design of Low-Carbon Multi-Energy Systems". In: *Applied Energy* 274, p. 115197. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2020.115197](https://doi.org/10.1016/j.apenergy.2020.115197). URL: <https://www.sciencedirect.com/science/article/pii/S0306261920307091>.
- Pfenninger, Stefan, Adam Hawkes, and James Keirstead (May 2014). "Energy Systems Modeling for Twenty-First Century Energy Challenges". In: *Renewable and Sustainable Energy Reviews* 33, pp. 74–86. ISSN: 1364-0321. doi: [10.1016/j.rser.2014.02.003](https://doi.org/10.1016/j.rser.2014.02.003). URL: <https://www.sciencedirect.com/science/article/pii/S1364032114000872>.
- Pfenninger, Stefan and Bryn Pickering (Sept. 2018). "Calliope: A Multi-Scale Energy Systems Modelling Framework". In: *Journal of Open Source Software* 3.29, p. 825. ISSN: 2475-9066. doi: [10.21105/joss.00825](https://doi.org/10.21105/joss.00825). URL: <http://joss.theoj.org/papers/10.21105/joss.00825>.
- Pfenninger, Stefan and Iain Staffell (Nov. 2016). "Long-Term Patterns of European PV Output Using 30 Years of Validated Hourly Reanalysis and Satellite Data". In: *Energy* 114, pp. 1251–1265. ISSN: 0360-5442. doi: [10.1016/j.energy.2016.08.060](https://doi.org/10.1016/j.energy.2016.08.060). URL: <https://www.sciencedirect.com/science/article/pii/S0360544216311744>.
- Phu, Pham Van, Truong Hoang Bao Huy, Seongkeun Park, and Daehee Kim (June 2024). "An IGDT Approach for the Multi-Objective Framework of Integrated Energy Hub with Renewable Energy Sources, Hybrid Energy Storage Systems, and Biomass-to-Hydrogen Technology". In: *Journal of Energy Storage* 89, p. 111488. ISSN: 2352-152X. doi: [10.1016/j.est.2024.111488](https://doi.org/10.1016/j.est.2024.111488). URL: <https://www.sciencedirect.com/science/article/pii/S2352152X24010739>.
- Pickering, Bryn (May 2019). "Practical Optimisation of District Energy Systems: Representation of Technology Characteristics, Demand Uncertainty, and System Robustness". In: doi: [10.17863/CAM.39849](https://doi.org/10.17863/CAM.39849). URL: <https://www.repository.cam.ac.uk/handle/1810/292696>.
- Pickering, Bryn and Ruchi Choudhary (Feb. 2019). "District Energy System Optimisation under Uncertain Demand: Handling Data-Driven Stochastic Profiles". In: *Applied Energy* 236, pp. 1138–1157. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2018.12.037](https://doi.org/10.1016/j.apenergy.2018.12.037). URL: <https://www.sciencedirect.com/science/article/pii/S0306261918318555>.
- Pickering, Bryn, Shintaro Ikeda, Ruchi Choudhary, and Ryozo Ooka (May 2016). *Comparison of Metaheuristic and Linear Programming Models for the Purpose of Optimising Building Energy Supply Operation Schedule*.

- Prataviera, Enrico, Jacopo Vivian, Giulia Lombardo, and Angelo Zarrella (Apr. 2022). "Evaluation of the Impact of Input Uncertainty on Urban Building Energy Simulations Using Uncertainty and Sensitivity Analysis". In: *Applied Energy* 311, p. 118691. issn: 0306-2619. doi: [10.1016/j.apenerg.2022.118691](https://doi.org/10.1016/j.apenerg.2022.118691). URL: <https://www.sciencedirect.com/science/article/pii/S0306261922001568>.
- Project - H2-Fifty (2022). URL: <https://www.h2-fifty.com/about/>.
- Rady, Mohammed, Mohd Khairul Idlan Muhammad, and Shamsuddin Shahid (Jan. 2025). "Evolving Typical Meteorological Year (TMY) Data for Building Energy Simulation: A Comprehensive Review of Methods, Challenges, and Future Directions". In: *Advances in Building Energy Research* 0.0, pp. 1–31. issn: 1751-2549. doi: [10.1080/17512549.2025.2457649](https://doi.org/10.1080/17512549.2025.2457649). URL: <https://doi.org/10.1080/17512549.2025.2457649>.
- Rahman, Aowabin, Vivek Sri Kumar, and Amanda D. Smith (Feb. 2018). "Predicting Electricity Consumption for Commercial and Residential Buildings Using Deep Recurrent Neural Networks". In: *Applied Energy* 212, pp. 372–385. issn: 0306-2619. doi: [10.1016/j.apenergy.2017.12.051](https://doi.org/10.1016/j.apenergy.2017.12.051). URL: <https://www.sciencedirect.com/science/article/pii/S0306261917317658>.
- Raiffa, Howard and Robert Schlaifer (1961). *Applied Statistical Decision Theory*. John Wiley & Sons. ISBN: 978-0-471-38349-9.
- Rainforth, Tom, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith (Feb. 2024). "Modern Bayesian Experimental Design". In: *Statistical Science* 39.1, pp. 100–114. issn: 0883-4237, 2168-8745. doi: [10.1214/23-STS915](https://projecteuclid.org/journals/statistical-science/volume-39/issue-1/Modern-Bayesian-Experimental-Design/10.1214/23-STS915.full). URL: <https://projecteuclid.org/journals/statistical-science/volume-39/issue-1/Modern-Bayesian-Experimental-Design/10.1214/23-STS915.full>.
- Raisch, Fabian, Thomas Krug, Christoph Goebel, and Benjamin Tischler (June 2025). "GenTL: A General Transfer Learning Model for Building Thermal Dynamics". In: *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. Rotterdam Netherlands: ACM, pp. 322–333. doi: [10.1145/3679240.3734589](https://doi.org/10.1145/3679240.3734589). URL: <https://dl.acm.org/doi/10.1145/3679240.3734589>.
- Raisch, Fabian, Max Langtry, Felix Koch, et al. (Aug. 2025). *Adapting to Change: A Comparison of Continual and Transfer Learning for Modeling Building Thermal Dynamics under Concept Drifts*. (Submitted to Energy and Buildings). doi: [10.48550/arXiv.2508.21615](https://doi.org/10.48550/arXiv.2508.21615). arXiv: [2508.21615 \[eess\]](https://arxiv.org/abs/2508.21615). URL: [http://arxiv.org/abs/2508.21615](https://arxiv.org/abs/2508.21615).
- Ramasamy, Vignesh, Jarett Zuboy, Michael Woodhouse, et al. (Sept. 2023). "U.S. Solar Photovoltaic System and Energy Storage Cost Benchmarks, With Minimum Sustainable Price Analysis: Q1 2023". In: *National Renewable Energy Laboratory*. URL: <https://www.nrel.gov/docs/fy23osti/87303.pdf>.
- Rentals, Sunbelt (Dec. 2023). *Thermtest TLS-100 Thermal Conductivity Meter*. Tech. rep. URL: <https://web.archive.org/web/20231206070141/https://www.inlec.com/thermtest-tls-100>.
- Risch, Stanley, Peter Remmen, and Dirk Müller (Jan. 2021). "Influence of Data Acquisition on the Bayesian Calibration of Urban Building Energy Models". In: *Energy and Buildings* 230, p. 110512. issn: 0378-7788. doi: [10.1016/j.enbuild.2020.110512](https://doi.org/10.1016/j.enbuild.2020.110512). URL: <https://www.sciencedirect.com/science/article/pii/S0378778820313748>.
- Rivalin, Lisa, Pascal Stabat, Dominique Marchio, Marcello Caciolo, and Frédéric Hopquin (May 2018). "A Comparison of Methods for Uncertainty and Sensitivity Analysis Applied to the Energy Performance of New Commercial Buildings". In: *Energy and Buildings* 166, pp. 489–504. issn: 0378-7788. doi: [10.1016/j.enbuild.2018.02.021](https://doi.org/10.1016/j.enbuild.2018.02.021). URL: <https://www.sciencedirect.com/science/article/pii/S0378778818305243>.

- Rockafellar, R. Tyrrell and Stanislav Uryasev (2000). "Optimization of Conditional Value-at-Risk". In: *The Journal of Risk* 2.3, pp. 21–41. ISSN: 14651211. doi: [10.21314/JOR.2000.038](https://doi.org/10.21314/JOR.2000.038). URL: <http://www.risk.net/journal-of-risk/technical-paper/2161159/optimization-conditional-value-risk>.
- Rysanek, Adam Martin and Ruchi Choudhary (Feb. 2013). "Optimum Building Energy Retrofits under Technical and Economic Uncertainty". In: *Energy and Buildings* 57, pp. 324–337. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2012.10.027](https://doi.org/10.1016/j.enbuild.2012.10.027). URL: <https://www.sciencedirect.com/science/article/pii/S0378778812005361>.
- Salinas, David, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski (July 2020). "DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks". In: *International Journal of Forecasting* 36.3, pp. 1181–1191. ISSN: 0169-2070. doi: [10.1016/j.ijforecast.2019.07.001](https://doi.org/10.1016/j.ijforecast.2019.07.001). URL: <https://www.sciencedirect.com/science/article/pii/S0169207019301888>.
- Salvador, Michaël and Stéphane Grieu (Apr. 2012). "Methodology for the Design of Energy Production and Storage Systems in Buildings: Minimization of the Energy Impact on the Electricity Grid". In: *Energy and Buildings* 47, pp. 659–673. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2012.01.006](https://doi.org/10.1016/j.enbuild.2012.01.006). URL: <https://www.sciencedirect.com/science/article/pii/S0378778812000084>.
- Sanajaoba Singh, Sarangthem and Eugene Fernandez (Jan. 2018). "Modeling, Size Optimization and Sensitivity Analysis of a Remote Hybrid Renewable Energy System". In: *Energy* 143, pp. 719–731. ISSN: 0360-5442. doi: [10.1016/j.energy.2017.11.053](https://doi.org/10.1016/j.energy.2017.11.053). URL: <https://www.sciencedirect.com/science/article/pii/S0360544217319023>.
- Savadkoohi, Marjan, Marcel Macarulla, and Miquel Casals (Jan. 2023). "Facilitating the Implementation of Neural Network-Based Predictive Control to Optimize Building Heating Operation". In: *Energy* 263, p. 125703. ISSN: 0360-5442. doi: [10.1016/j.energy.2022.125703](https://doi.org/10.1016/j.energy.2022.125703). URL: <https://www.sciencedirect.com/science/article/pii/S0360544222025890>.
- Schwarzlander, Harry (Feb. 2011). *Probability Concepts and Theory for Engineers*. John Wiley & Sons. ISBN: 978-0-470-74855-8.
- Semeraro, Concetta, Mario Lezoche, Hervé Panetto, and Michele Dassisti (Sept. 2021). "Digital Twin Paradigm: A Systematic Literature Review". In: *Computers in Industry* 130, p. 103469. ISSN: 0166-3615. doi: [10.1016/j.compind.2021.103469](https://doi.org/10.1016/j.compind.2021.103469). URL: <https://www.sciencedirect.com/science/article/pii/S0166361521000762>.
- Sepulveda, Nestor A., Jesse D. Jenkins, Aurora Edington, Dharik S. Mallapragada, and Richard K. Lester (May 2021). "The Design Space for Long-Duration Energy Storage in Decarbonized Power Systems". In: *Nature Energy* 6.5, pp. 506–516. ISSN: 2058-7546. doi: [10.1038/s41560-021-00796-8](https://doi.org/10.1038/s41560-021-00796-8). URL: <https://www.nature.com/articles/s41560-021-00796-8>.
- Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas (Jan. 2016). "Taking the Human Out of the Loop: A Review of Bayesian Optimization". In: *Proceedings of the IEEE* 104.1, pp. 148–175. ISSN: 1558-2256. doi: [10.1109/JPROC.2015.2494218](https://doi.org/10.1109/JPROC.2015.2494218). URL: <https://ieeexplore.ieee.org/abstract/document/7352306>.
- Sharma, Pooja, Mohan Kolhe, and Arvind Sharma (Jan. 2020). "Economic Performance Assessment of Building Integrated Photovoltaic System with Battery Energy Storage under Grid Constraints". In: *Renewable Energy* 145, pp. 1901–1909. ISSN: 0960-1481. doi: [10.1016/j.renene.2019.07.099](https://doi.org/10.1016/j.renene.2019.07.099). URL: <https://www.sciencedirect.com/science/article/pii/S096014811931119X>.
- Smith, C. a. B. (July 1945). "Theory of Games and Economic Behaviour. By John Von Neumann and Oskar Morgenstern. Pp. Xviii, 625. 66s. 6d. 1944. (Princeton U.P.; Humphrey Milford)". In: *The Mathematical Gazette* 29.285, pp. 131–133. ISSN: 0025-5572, 2056-6328. doi: [10.2307/3610940](https://doi.org/10.2307/3610940).

- URL: <https://www.cambridge.org/core/journals/mathematical-gazette/article/abs/theory-of-games-and-economic-behaviour-by-john-von-neumann-and-oskar-morgenstern-pp-xviii-625-66s-6d-1944-princeton-up-humphrey-milford/76E26E70D0D9B6AF26DB67A6A10D58CF>.
- Souto, Laiz (Oct. 2018). "Scenario-Based Decentralized MPC Framework for Power Distribution Grids with Self-Interested Energy Prosumers". In: *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6. doi: 10.1109/ISGETurope.2018.8571459. URL: <https://ieeexplore.ieee.org/abstract/document/8571459>.
- Spitler, J D, C Yavuzturk, and S J Rees (2000). "In Situ Measurement of Ground Thermal Properties". In: *Proceedings of Terrastock 2000*, pp. 165–170. URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3856e0aadf680f5704e60f6a6c0aea32ac674e08>.
- Spivey, W. Allen (1968). "Decision Making and Probabilistic Programming". In: *IMR* 9.2, p. 57. ISSN: 08848211. URL: <https://www.proquest.com/docview/214191669/citation/33C3557594244378/PQ/1>.
- Staffell, Iain and Stefan Pfenninger (2016a). *Renewables.Ninja*. URL: <https://www.renewables.ninja/>.
- (Nov. 2016b). "Using Bias-Corrected Reanalysis to Simulate Current and Future Wind Power Output". In: *Energy* 114, pp. 1224–1239. ISSN: 0360-5442. doi: 10.1016/j.energy.2016.08.068. URL: <https://www.sciencedirect.com/science/article/pii/S0360544216311811>.
- Stehly, Tyler, Patrick Duffy, and Daniel Mulas Hernando (Dec. 2023). "2022 Cost of Wind Energy Review". In: *National Renewable Energy Laboratory*. URL: <https://www.nrel.gov/docs/fy24osti/88335.pdf>.
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum (Apr. 2020). "Energy and Policy Considerations for Modern Deep Learning Research". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.09, pp. 13693–13696. ISSN: 2374-3468. doi: 10.1609/aaai.v34i09.7123. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/7123>.
- Sun, Ying, Fariborz Haghhighat, and Benjamin C. M. Fung (Aug. 2020a). "A Review of The-State-of-the-Art in Data-Driven Approaches for Building Energy Prediction". In: *Energy and Buildings* 221, p. 110022. ISSN: 0378-7788. doi: 10.1016/j.enbuild.2020.110022. URL: <https://www.sciencedirect.com/science/article/pii/S0378778819339313>.
- (Aug. 2020b). "A Review of The-State-of-the-Art in Data-Driven Approaches for Building Energy Prediction". In: *Energy and Buildings* 221, p. 110022. ISSN: 0378-7788. doi: 10.1016/j.enbuild.2020.110022. URL: <https://www.sciencedirect.com/science/article/pii/S0378778819339313>.
- Sun, Yongjun (Jan. 2015). "Sensitivity Analysis of Macro-Parameters in the System Design of Net Zero Energy Building". In: *Energy and Buildings* 86, pp. 464–477. ISSN: 0378-7788. doi: 10.1016/j.enbuild.2014.10.031. URL: <https://www.sciencedirect.com/science/article/pii/S0378778814008767>.
- Szaniawski, Klemens (1967). "The Value of Perfect Information". In: *Synthese* 17.4, pp. 408–424. ISSN: 0039-7857. JSTOR: 20114577. URL: <https://www.jstor.org/stable/20114577>.
- Tang, Fujiao and Hossein Nowamooz (Dec. 2019). "Sensitive Analysis on the Effective Soil Thermal Conductivity of the Thermal Response Test Considering Various Testing Times, Field Conditions and U-pipe Lengths". In: *Renewable Energy* 143, pp. 1732–1743. ISSN: 0960-1481. doi: 10.1016/j.renene.2019.05.120. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119307931>.

- Terry, Nicola and Ray Galvin (Aug. 2023). "How Do Heat Demand and Energy Consumption Change When Households Transition from Gas Boilers to Heat Pumps in the UK". In: *Energy and Buildings* 292, p. 113183. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2023.113183](https://doi.org/10.1016/j.enbuild.2023.113183). URL: <https://www.sciencedirect.com/science/article/pii/S0378778823004139>.
- Thieblemont, Hélène, Fariborz Haghishat, Ryozo Ooka, and Alain Moreau (Oct. 2017). "Predictive Control Strategies Based on Weather Forecast in Buildings with Energy Storage System: A Review of the State-of-the Art". In: *Energy and Buildings* 153, pp. 485–500. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2017.08.010](https://doi.org/10.1016/j.enbuild.2017.08.010). URL: <https://www.sciencedirect.com/science/article/pii/S0378778816320333>.
- Tian, Wei (Apr. 2013). "A Review of Sensitivity Analysis Methods in Building Energy Analysis". In: *Renewable and Sustainable Energy Reviews* 20, pp. 411–419. ISSN: 1364-0321. doi: [10.1016/j.rser.2012.12.014](https://doi.org/10.1016/j.rser.2012.12.014). URL: <https://www.sciencedirect.com/science/article/pii/S1364032112007101>.
- Tian, Wei, Yeonsook Heo, Pieter Wilde, et al. (Oct. 2018). "A Review of Uncertainty Analysis in Building Energy Assessment". In: *Renewable and Sustainable Energy Reviews* 93, pp. 285–301. ISSN: 1364-0321. doi: [10.1016/j.rser.2018.05.029](https://doi.org/10.1016/j.rser.2018.05.029). URL: <https://www.sciencedirect.com/science/article/pii/S136403211830368X>.
- Tian, Wei, Song Yang, Zhanyong Li, et al. (May 2016). "Identifying Informative Energy Data in Bayesian Calibration of Building Energy Models". In: *Energy and Buildings* 119, pp. 363–376. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2016.03.042](https://doi.org/10.1016/j.enbuild.2016.03.042). URL: <https://www.sciencedirect.com/science/article/pii/S0378778816301967>.
- Tostado-Vélez, Marcos, Ahmad Rezaee Jordehi, Seyed Amir Mansouri, et al. (Aug. 2024). "Risk-Averse Electrolyser Sizing in Industrial Parks: An Efficient Stochastic-Robust Approach". In: *Applied Energy* 367, p. 123389. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2024.123389](https://doi.org/10.1016/j.apenergy.2024.123389). URL: <https://www.sciencedirect.com/science/article/pii/S0306261924007724>.
- Touretzky, Cara R. and Michael Baldea (Aug. 2014). "Integrating Scheduling and Control for Economic MPC of Buildings with Energy Storage". In: *Journal of Process Control. Economic Nonlinear Model Predictive Control* 24.8, pp. 1292–1300. ISSN: 0959-1524. doi: [10.1016/j.jprocont.2014.04.015](https://doi.org/10.1016/j.jprocont.2014.04.015). URL: <https://www.sciencedirect.com/science/article/pii/S0959152414001164>.
- Tröndle, Tim and Ruchi Choudhary (2017). "Occupancy Based Thermal Energy Modelling in the Urban Residential Sector". In: *WIT Transactions on Ecology and the Environment* 224, pp. 31–44.
- Tuffaha, Haitham W., Gordon Louisa G., and Paul A. and Scuffham (June 2014). "Value of Information Analysis in Healthcare: A Review of Principles and Applications". In: *Journal of Medical Economics* 17.6, pp. 377–383. ISSN: 1369-6998. doi: [10.3111/13696998.2014.907170](https://doi.org/10.3111/13696998.2014.907170). URL: <https://doi.org/10.3111/13696998.2014.907170>.
- Tumminia, G., F. Guarino, S. Longo, et al. (Jan. 2020). "Grid Interaction and Environmental Impact of a Net Zero Energy Building". In: *Energy Conversion and Management* 203, p. 112228. ISSN: 0196-8904. doi: [10.1016/j.enconman.2019.112228](https://doi.org/10.1016/j.enconman.2019.112228). URL: <https://www.sciencedirect.com/science/article/pii/S0196890419312348>.
- van de Schoot, Rens, Sarah Depaoli, Ruth King, et al. (Jan. 2021). "Bayesian Statistics and Modelling". In: *Nature Reviews Methods Primers* 1.1, p. 1. ISSN: 2662-8449. doi: [10.1038/s43586-020-00001-2](https://doi.org/10.1038/s43586-020-00001-2). URL: <https://doi.org/10.1038/s43586-020-00001-2>.
- Vazquez-Canteli, Jose R., Sourav Dey, Gregor Henze, and Zoltan Nagy (Dec. 2020). *CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and*

- Urban Energy Management*. doi: [10.48550/arXiv.2012.10504](https://doi.org/10.48550/arXiv.2012.10504). arXiv: [2012.10504 \[cs\]](https://arxiv.org/abs/2012.10504). URL: <http://arxiv.org/abs/2012.10504>.
- Vazquez-Canteli, Jose R., Gregor Henze, and Zoltan Nagy (Nov. 2020). "MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings". In: *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. Virtual Event Japan: ACM, pp. 170–179. ISBN: 978-1-4503-8061-4. doi: [10.1145/3408308.3427604](https://doi.org/10.1145/3408308.3427604). URL: <https://dl.acm.org/g/doi/10.1145/3408308.3427604>.
- Vázquez-Canteli, José R., Jérôme Kämpf, Gregor Henze, and Zoltan Nagy (Nov. 2019). "CityLearn v1.0: An OpenAI Gym Environment for Demand Response with Deep Reinforcement Learning". In: *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. New York NY USA: ACM, pp. 356–357. ISBN: 978-1-4503-7005-9. doi: [10.1145/3360322.3360998](https://doi.org/10.1145/3360322.3360998). URL: <https://dl.acm.org/doi/10.1145/3360322.3360998>.
- Verleysen, Michel and Damien François (2005). "The Curse of Dimensionality in Data Mining and Time Series Prediction". In: *Computational Intelligence and Bioinspired Systems*. Ed. by Joan Cabestany, Alberto Prieto, and Francisco Sandoval. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 758–770. ISBN: 978-3-540-32106-4. doi: [10.1007/11494669_93](https://doi.org/10.1007/11494669_93).
- Wagg, David J., Christopher Burr, Jason Shepherd, et al. (Jan. 2025). "The Philosophical Foundations of Digital Twinning". In: *Data-Centric Engineering* 6, e12. ISSN: 2632-6736. doi: [10.1017/dce.2025.4](https://doi.org/10.1017/dce.2025.4). URL: <https://www.cambridge.org/core/journals/data-centric-engineering/article/philoosophical-foundations-of-digital-twinning/7C2E25F20DDB7EF7DDD560363DE9E171#>.
- Wang, Chao, Martina Ferrando, Francesco Causone, et al. (June 2022). "Data Acquisition for Urban Building Energy Modeling: A Review". In: *Building and Environment* 217, p. 109056. ISSN: 0360-1323. doi: [10.1016/j.buildenv.2022.109056](https://doi.org/10.1016/j.buildenv.2022.109056). URL: <https://www.sciencedirect.com/science/article/pii/S0360132322002955>.
- Wang, Chengshan, Bingqi Jiao, Li Guo, et al. (Apr. 2016). "Robust Scheduling of Building Energy System under Uncertainty". In: *Applied Energy* 167, pp. 366–376. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2015.09.070](https://doi.org/10.1016/j.apenergy.2015.09.070). URL: <https://www.sciencedirect.com/science/article/pii/S0306261915011794>.
- Wang, Jinshi, Kai Xue, Yanjun Guo, et al. (Sept. 2022). "Multi-Objective Capacity Programming and Operation Optimization of an Integrated Energy System Considering Hydrogen Energy Storage for Collective Energy Communities". In: *Energy Conversion and Management* 268, p. 116057. ISSN: 0196-8904. doi: [10.1016/j.enconman.2022.116057](https://doi.org/10.1016/j.enconman.2022.116057). URL: <https://www.sciencedirect.com/science/article/pii/S0196890422008457>.
- Wang, Jun, Jianli Chen, and Yuqing Hu (Oct. 2022). "A Science Mapping Approach Based Review of Model Predictive Control for Smart Building Operation Management". In: *Journal of Civil Engineering and Management* 28.8, pp. 661–679. ISSN: 1822-3605. doi: [10.3846/jcem.2022.17566](https://doi.org/10.3846/jcem.2022.17566). URL: <https://jau.vgtu.lt/index.php/JCEM/article/view/17566>.
- Wang, Zhe and Tianzhen Hong (July 2020). "Reinforcement Learning for Building Controls: The Opportunities and Challenges". In: *Applied Energy* 269, p. 115036. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2020.115036](https://doi.org/10.1016/j.apenergy.2020.115036). URL: <https://www.sciencedirect.com/science/article/pii/S0306261920305481>.
- Ward, Rebecca (Sept. 2021). "A Data-Centric Stochastic Model for Simulation of Occupant-Related Energy Demand in Buildings". In: URL: <https://www.repository.cam.ac.uk/handle/1810/328479>.

- Ward, Rebecca and Ruchi Choudhary (2021). "A Tool for Generation of Stochastic Occupant-Based Internal Loads Using a Functional Data Analysis Approach to Re-Define 'Activity'". In: *Journal of Building Performance Simulation* 14.3, pp. 303–327.
- Ward, Rebecca, Ruchi Choudhary, Y. Heo, and J. A. D. Aston (Sept. 2019). "A Data-Centric Bottom-up Model for Generation of Stochastic Internal Load Profiles Based on Space-Use Type". In: *Journal of Building Performance Simulation* 12.5, pp. 620–636. ISSN: 1940-1493. doi: [10.1080/19401493.2019.1583287](https://doi.org/10.1080/19401493.2019.1583287). URL: <https://doi.org/10.1080/19401493.2019.1583287>.
- Wendling, Zachary A. (Feb. 2019). "Bridges beyond Renewable Energy: Decarbonizing the Global Electricity Sector under Uncertainty". In: *Energy Research & Social Science* 48, pp. 235–245. ISSN: 2214-6296. doi: [10.1016/j.erss.2018.09.020](https://doi.org/10.1016/j.erss.2018.09.020). URL: <https://www.sciencedirect.com/science/article/pii/S2214629618310387>.
- Willan, Andrew R. and Eleanor M. Pinto (2005). "The Value of Information and Optimal Clinical Trial Design". In: *Statistics in Medicine* 24.12, pp. 1791–1806. ISSN: 1097-0258. doi: [10.1002/sim.2069](https://doi.org/10.1002/sim.2069). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2069>.
- Williams, A. C. (Dec. 1965). "On Stochastic Linear Programming". In: *Journal of the Society for Industrial and Applied Mathematics* 13.4, pp. 927–940. ISSN: 0368-4245, 2168-3484. doi: [10.1137/0113060](https://doi.org/10.1137/0113060). URL: <http://pubs.siam.org/doi/10.1137/0113060>.
- Wilson, Robert (Aug. 1966). "On Programming under Uncertainty". In: *Operations Research* 14.4, pp. 652–657. ISSN: 0030-364X. doi: [10.1287/opre.14.4.652](https://doi.org/10.1287/opre.14.4.652). URL: <https://pubsonline.informs.org/doi/abs/10.1287/opre.14.4.652>.
- Winschermann, Leoni, Nataly Bañol Arias, Gerwin Hoogsteen, and Johann Hurink (Oct. 2023). "Assessing the Value of Information for Electric Vehicle Charging Strategies at Office Buildings". In: *Renewable and Sustainable Energy Reviews* 185, p. 113600. ISSN: 1364-0321. doi: [10.1016/j.rser.2023.113600](https://doi.org/10.1016/j.rser.2023.113600). URL: <https://www.sciencedirect.com/science/article/pii/S1364032123004574>.
- Wu, Yi, Jingjing An, Chenxi Gui, Chan Xiao, and Da Yan (June 2023). "A Global Typical Meteorological Year (TMY) Database on ERA5 Dataset". In: *Building Simulation* 16.6, pp. 1013–1026. ISSN: 1996-8744. doi: [10.1007/s12273-023-1015-3](https://doi.org/10.1007/s12273-023-1015-3). URL: <https://doi.org/10.1007/s12273-023-1015-3>.
- Xia, Jianjun, Tianzhen Hong, Qi Shen, et al. (Aug. 2014). "Comparison of Building Energy Use Data between the United States and China". In: *Energy and Buildings* 78, pp. 165–175. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2014.04.031](https://doi.org/10.1016/j.enbuild.2014.04.031). URL: <https://www.sciencedirect.com/science/article/pii/S0378778814003533>.
- Yamamoto, Shinya, Masahiro Furukakoi, Akie Uehara, et al. (Sept. 2024). "MPC-based Robust Optimization of Smart Apartment Building Considering Uncertainty for Conservative Reduction". In: *Energy and Buildings* 318, p. 114461. ISSN: 0378-7788. doi: [10.1016/j.enbuild.2024.114461](https://doi.org/10.1016/j.enbuild.2024.114461). URL: <https://www.sciencedirect.com/science/article/pii/S0378778824005772>.
- Yang, Yongbiao, Aoxue Cheng, Qingshan Xu, Yang Xu, and Xin Shen (May 2024). "Planning Microgrids in Industrial Parks Considering the Uncertainty of Carbon Price and Renewable Energy: A Stochastic Approach". In: *2024 IEEE 7th International Electrical and Energy Conference (CIEEC)*, pp. 117–122. doi: [10.1109/CIEEC60922.2024.10583205](https://doi.org/10.1109/CIEEC60922.2024.10583205). URL: <https://ieeexplore.ieee.org/abstract/document/10583205>.
- Yang, Yun, Shijie Zhang, and Yunhan Xiao (Oct. 2015). "An MILP (Mixed Integer Linear Programming) Model for Optimal Design of District-Scale Distributed Energy Resource Systems". In:

- Energy* 90, pp. 1901–1915. ISSN: 0360-5442. doi: [10.1016/j.energy.2015.07.013](https://doi.org/10.1016/j.energy.2015.07.013). URL: <https://www.sciencedirect.com/science/article/pii/S036054421500907X>.
- Yue, Xiufeng, Steve Pye, Joseph DeCarolis, et al. (Aug. 2018). “A Review of Approaches to Uncertainty Assessment in Energy System Optimization Models”. In: *Energy Strategy Reviews* 21, pp. 204–217. ISSN: 2211-467X. doi: [10.1016/j.esr.2018.06.003](https://doi.org/10.1016/j.esr.2018.06.003). URL: <https://www.sciencedirect.com/science/article/pii/S2211467X18300543>.
- Zeng, Ailing, Muxi Chen, Lei Zhang, and Qiang Xu (June 2023). “Are Transformers Effective for Time Series Forecasting?” In: *Proceedings of the AAAI Conference on Artificial Intelligence* 37.9, pp. 11121–11128. ISSN: 2374-3468. doi: [10.1609/aaai.v37i9.26317](https://doi.org/10.1609/aaai.v37i9.26317). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/26317>.
- Zhai, Zhiqiang and Andrea Salazar (Jan. 2020). “Assessing the Implications of Submetering with Energy Analytics to Building Energy Savings”. In: *Energy and Built Environment* 1.1, pp. 27–35. ISSN: 2666-1233. doi: [10.1016/j.enbenv.2019.08.002](https://doi.org/10.1016/j.enbenv.2019.08.002). URL: <https://www.sciencedirect.com/science/article/pii/S2666123319300029>.
- Zhan, Sicheng and Adrian Chong (May 2021). “Data Requirements and Performance Evaluation of Model Predictive Control in Buildings: A Modeling Perspective”. In: *Renewable and Sustainable Energy Reviews* 142, p. 110835. ISSN: 1364-0321. doi: [10.1016/j.rser.2021.110835](https://doi.org/10.1016/j.rser.2021.110835). URL: <https://www.sciencedirect.com/science/article/pii/S1364032121001295>.
- Zhan, Sicheng, Yue Lei, and Adrian Chong (2023). “Comparing Model Predictive Control and Reinforcement Learning for the Optimal Operation of Building-PV-battery Systems”. In: *E3S Web of Conferences* 396, p. 04018. ISSN: 2267-1242. doi: [10.1051/e3sconf/202339604018](https://doi.org/10.1051/e3sconf/202339604018). URL: https://www.e3s-conferences.org/articles/e3sconf/abs/2023/33/e3sconf_iqvec2023_04018/e3sconf_iqvec2023_04018.html.
- Zhan, Sicheng, Yue Lei, Yuan Jin, Da Yan, and Adrian Chong (Oct. 2022). “Impact of Occupant Related Data on Identification and Model Predictive Control for Buildings”. In: *Applied Energy* 323, p. 119580. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2022.119580](https://doi.org/10.1016/j.apenergy.2022.119580). URL: <https://www.sciencedirect.com/science/article/pii/S0306261922008881>.
- Zhan, Sicheng, Matias Quintana, Clayton Miller, and Adrian Chong (Dec. 2022). “From Model-Centric to Data-Centric: A Practical MPC Implementation Framework for Buildings”. In: *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. BuildSys ’22. New York, NY, USA: Association for Computing Machinery, pp. 270–273. ISBN: 978-1-4503-9890-9. doi: [10.1145/3563357.3564077](https://doi.org/10.1145/3563357.3564077). URL: <https://dl.acm.org/g/doi/10.1145/3563357.3564077>.
- Zhang, Hao, Yajie Zou, Xiaoxue Yang, and Hang Yang (Aug. 2022). “A Temporal Fusion Transformer for Short-Term Freeway Traffic Speed Multistep Prediction”. In: *Neurocomputing* 500, pp. 329–340. ISSN: 0925-2312. doi: [10.1016/j.neucom.2022.05.083](https://doi.org/10.1016/j.neucom.2022.05.083). URL: <https://www.sciencedirect.com/science/article/pii/S092523122200666X>.
- Zhang, Huiliang, Sayani Seal, Di Wu, et al. (June 2021). *Data-Driven Model Predictive and Reinforcement Learning Based Control for Building Energy Management: A Survey*. doi: [10.48550/arXiv.2106.14450](https://arxiv.org/abs/2106.14450). arXiv: [2106.14450 \[eess\]](https://arxiv.org/abs/2106.14450). URL: [http://arxiv.org/abs/2106.14450](https://arxiv.org/abs/2106.14450).
- Zhang, Liang, Jin Wen, Yanfei Li, et al. (Mar. 2021). “A Review of Machine Learning in Building Load Prediction”. In: *Applied Energy* 285, p. 116452. ISSN: 0306-2619. doi: [10.1016/j.apenergy.2021.116452](https://doi.org/10.1016/j.apenergy.2021.116452). URL: <https://www.sciencedirect.com/science/article/pii/S0306261921000209>.

- Zhang, Meng, Michael-Allan Millar, Zhibin Yu, and James Yu (Nov. 2022). "An Assessment of the Impacts of Heat Electrification on the Electric Grid in the UK". In: *Energy Reports* 8, pp. 14934–14946. ISSN: 2352-4847. DOI: [10.1016/j.egyr.2022.10.408](https://doi.org/10.1016/j.egyr.2022.10.408). URL: <https://www.sciencedirect.com/science/article/pii/S2352484722023393>.
- Zhang, Qingang, Yunqi Huang, Chin-Boon Chng, Chee-Kong Chui, and Poh-Seng Lee (Sept. 2023). "Investigations on Machine Learning-Based Control-Oriented Modeling Using Historical Thermal Data of Buildings". In: *Building and Environment* 243, p. 110595. ISSN: 0360-1323. DOI: [10.1016/j.buildenv.2023.110595](https://doi.org/10.1016/j.buildenv.2023.110595). URL: <https://www.sciencedirect.com/science/article/pii/S0360132323006224>.
- Zhang, Tairan and Behrouz Sobhani (May 2023). "Optimal Economic Programming of an Energy Hub in the Power System While Taking into Account the Uncertainty of Renewable Resources, Risk-Taking and Electric Vehicles Using a Developed Routing Method". In: *Energy* 271, p. 126938. ISSN: 0360-5442. DOI: [10.1016/j.energy.2023.126938](https://doi.org/10.1016/j.energy.2023.126938). URL: <https://www.sciencedirect.com/science/article/pii/S0360544223003328>.
- Zhang, Wei-Heng, Da-Gang Lu, Jianjun Qin, Sebastian Thöns, and Michael Havbro Faber (May 2021). "Value of Information Analysis in Civil and Infrastructure Engineering: A Review". In: *Journal of Infrastructure Preservation and Resilience* 2.1, p. 16. ISSN: 2662-2521. DOI: [10.1186/s43065-021-00027-0](https://doi.org/10.1186/s43065-021-00027-0). URL: <https://doi.org/10.1186/s43065-021-00027-0>.
- Zhang, Yumeng, Ligang Wang, Ningling Wang, et al. (Dec. 2019). "Balancing Wind-Power Fluctuation via Onsite Storage under Uncertainty: Power-to-hydrogen-to-power versus Lithium Battery". In: *Renewable and Sustainable Energy Reviews* 116, p. 109465. ISSN: 1364-0321. DOI: [10.1016/j.rser.2019.109465](https://doi.org/10.1016/j.rser.2019.109465). URL: <https://www.sciencedirect.com/science/article/pii/S1364032119306732>.
- Zhao, Kaiguang, Michael A. Wulder, Tongxi Hu, et al. (Oct. 2019). "Detecting Change-Point, Trend, and Seasonality in Satellite Time Series Data to Track Abrupt Changes and Nonlinear Dynamics: A Bayesian Ensemble Algorithm". In: *Remote Sensing of Environment* 232, p. 111181. ISSN: 0034-4257. DOI: [10.1016/j.rse.2019.04.034](https://doi.org/10.1016/j.rse.2019.04.034). URL: <https://www.sciencedirect.com/science/article/pii/S0034425719301853>.
- Zhou, Yuan, Jiangjiang Wang, Yi Liu, Rujing Yan, and Yanpeng Ma (Oct. 2021). "Incorporating Deep Learning of Load Predictions to Enhance the Optimal Active Energy Management of Combined Cooling, Heating and Power System". In: *Energy* 233, p. 121134. ISSN: 0360-5442. DOI: [10.1016/j.energy.2021.121134](https://doi.org/10.1016/j.energy.2021.121134). URL: <https://www.sciencedirect.com/science/article/pii/S0360544221013827>.
- Zhu, Li, Jiqiang Zhang, Yuzhe Gao, et al. (Jan. 2022). "Uncertainty and Sensitivity Analysis of Cooling and Heating Loads for Building Energy Planning". In: *Journal of Building Engineering* 45, p. 103440. ISSN: 2352-7102. DOI: [10.1016/j.jobe.2021.103440](https://doi.org/10.1016/j.jobe.2021.103440). URL: <https://www.sciencedirect.com/science/article/pii/S2352710221012985>.
- Zhu, Xiaoyu, Xingxing Zhang, Pu Gong, and Yu Li (Aug. 2023). "A Review of Distributed Energy System Optimization for Building Decarbonization". In: *Journal of Building Engineering* 73, p. 106735. ISSN: 2352-7102. DOI: [10.1016/j.jobe.2023.106735](https://doi.org/10.1016/j.jobe.2023.106735). URL: <https://www.sciencedirect.com/science/article/pii/S2352710223009142>.
- Zhuang, Chaoqun, Ruchi Choudhary, and Anna Mavrogianni (July 2023). "Uncertainty-Based Optimal Energy Retrofit Methodology for Building Heat Electrification with Enhanced Energy Flexibility and Climate Adaptability". In: *Applied Energy* 341, p. 121111. ISSN: 0306-2619. DOI: [10.1016/j.apenergy.2023.121111](https://doi.org/10.1016/j.apenergy.2023.121111). URL: <https://www.sciencedirect.com/science/article/pii/S0306261923004750>.