LCA-based Semantically-enabled Framework for the Dynamic Optimisation of a Building's Energy and Environmental Performance

Abdulrahman Fnais

A thesis submitted in partial fulfilment of the requirement of the degree of Doctor of Philosophy

School of Engineering

Cardiff University

April 2023

Abstract

The rapid growth of cities, and the high energy consumption and Greenhouse Gases (GHGs) emissions of the buildings are significant challenges to reducing the environmental impact of the built environment. Life Cycle Assessment (LCA) can help to address these challenges by assessing the full life cycle of buildings and identifying areas for improvement. However, the complexity of the building sector, including variations in building usage, energy supply, and regulations, makes it difficult to consistently apply LCA methodologies to buildings. Thus, a different approach or methodology is required to improve the adoption and streamline the application of LCA in the building domain.

This thesis presents a comprehensive approach to facilitate the application of LCA in buildings, with a focus on enhancing their energy and environmental performance. A framework was developed to overcome the limitations of current LCA solutions and provide a comprehensive approach to explore various scenarios during the operation of buildings by integrating various domain models and data sources. This study demonstrated the practical application of the developed framework in addressing the research questions through a specific use case. The use case showed how the framework that was developed during this work could be applied to address the challenges of reducing the environmental impacts of building energy consumption during the operation phase. The proposed optimisation strategy for mechanical ventilation systems using genetic algorithms, coupled with machine learning models, provided a practical solution that minimises energy consumption while ensuring that indoor CO_2 levels remain within acceptable limits.

Finally, a lightweight ontology was developed for semantically-enabled LCA in buildings. The ontology was created by identifying the domain concepts and the relationships between the identified concepts. The ontology schema was developed using a modular approach, with three interconnected modules: the Observation module, Service module, and Building module. The ontology was evaluated through SPARQL queries and was effective in providing answers to questions from various domains. The developed ontology highlights the importance of leveraging semantics to integrate information and data from different sources, facilitating the application of LCA, and enhancing interoperability and information exchange across domains.

Acknowledgement

I would like to express my deepest gratitude to my main supervisor, Professor Yacine Rezgui, for his exceptional guidance, unwavering support, and invaluable mentorship throughout my PhD research. I am grateful to have had the privilege of working under his supervision.

I would also like to express my sincere gratitude to my second supervisor, Dr. Ioan Petri, for his invaluable support and insights throughout my research.

I would also like to extend my gratitude to the amazing team from the SemanticLCA project for their collaboration, and generous sharing of knowledge and expertise.

I am particularly grateful to Dr. Ali Ghoroghi for his support to my work, and the countless hours that we spent brainstorming and discussing various research ideas. Our engaging conversations, which sometimes strayed from my research topic to include shared interests in history and politics, brought some much-needed enjoyment to the process.

I am also deeply grateful to my parents and siblings, for their unwavering encouragement and support. Their belief in me and constant prayers were a source of strength and motivation throughout my journey.

I am humbled by the love and steadfast support of my wife Maram and daughter Lana during the years of living abroad to pursue my postgraduate degrees. They have been my constant companions, providing me with the strength and encouragement that I needed to keep going. I am truly grateful to my wife, a true friend and confidante, who offered me solace in moments of despair.

List of Publications

1. Fnais, Abdulrahman, et al. "The application of life cycle assessment in buildings: challenges, and directions for future research." The International Journal of Life Cycle Assessment 27.5 (2022): 627-654.
2. Fnais, Abdulrahman, et al. "SemanticLCA: A new generation of life cycle assessment methods applied to buildings." 2022 IEEE 28th International Conference on Engineering, Technology and Innovation (ICE/ITMC) 31st International Association For Management of Technology (IAMOT) Joint Conference. IEEE, 2022.

Contents

List of Figures

List of Tables

Nomenclature

LCA	Life Cycle Assessment
IoT	Internet of Things
GHGs	Greenhouse Gases
LCC	Life Cycle Costing
WWR	Window-to-Wall ratio
DEA	Data Envelopment Analysis
GWP	Global Warming Potential
EoL	End-of-Life
PEF	Product Environmental Footprint
ILCD	International Reference Life Cycle Data
LCI	Life Cycle Inventory
LCEA	Life Cycle Energy Assessment
CO_2	Carbon Dioxide
DLCA	Dynamic Life Cycle Assessment
ANN	Artificial Neural Networks
BIM	Building Information Modelling
EPDs	Environmental Product Declarations
ML	Machine-learning
GIS	Geographical Information Systems
LCIA	Life Cycle Impact Assessment
ICT	Information and Communication Technology
RF	Random Forest
GA	Genetic Algorithms
RDF	Resource Description Framework

OWL	Web Ontology Language
SPARQL	SPARQL Protocol and RDF Query Language

Chapter 1

Introduction

1.1 Background

Globally, the population who live in cities is predicted to grow to 68% by 2050 [1]. However, cities are currently responsible for 75% of global energy consumption and Greenhouse Gases (GHGs) emissions, with over 40% of total energy consumption attributed to buildings [2]. Moreover, the building sector is a key consumer of natural resources. In Europe, buildings are responsible for 33% of waste and 22% of hazardous waste production [3]. The special report on the impact of global warming of 1.5°C [4] was yet another call to implement measures to mitigate GHGs emissions and to devise new adaptation scenarios. In this context, Life Cycle Assessment (LCA) can help to quantify the environmental pressures, the trade-offs, and the areas to achieve improvements considering the full life cycle of built assets, from design to recycling. However, current approaches to LCA do not consistently factor in (both in the foreground and background inventory systems) life cycle variations in: (a) building usage, (b) energy supply (including from renewable sources), and (c) building and environmental regulations, as well as other changes over the building's or district's lifetime [5–7]. These include: (a) a change in the energy mix of a building or district, or upgrading/retrofitting the energy system(s) in place; and (b) time-increase of energy demand during the lifetime of a building due to a wide range of reasons, including changes in occupancy patterns.

LCA is an important instrument to help reduce the overall environmental burden of buildings. It can also provide insights into the upstream and downstream trade-offs that are associated with environmental pressures, health and well-being, and the consumption of natural resources. Therefore, LCA can inform policy making by providing valuable information on the environmental performance of built assets. However, the current LCA methods and tools have a number of limitations and challenges, including: (a) site-specific considerations [6], several local impacts need to be considered in building assessments (e.g., the microclimate); (b) model complexity [5], buildings involve a wide range of material and products that interact as part of a complex assembly or system; (c) scenario uncertainty [5, 6], the long use phase of buildings (including the potential for future renovation) poses uncertainty problems in LCA that are not currently addressed; (d) health and well-being [6, 7], traditional LCA methodologies do not address the indoor and outdoor environmental impacts of a building on health and well-being; (e) recycled material data [5, 8], there is a lack of data on the use of waste and recycled materials as new building materials; and (f) there is a lack of consideration of the social and economic aspects [5, 8].

1.2 LCA Across the Life-cycle Stages

This section provides a brief overview of the importance and challenges of applying LCA across key stages of a built asset's life cycle: design, retrofit and construction, operation, and End-of-Life (EoL) treatment.

In the design stage: There is an increasing demand for LCA modelling approaches that can be initiated during the early-design stage and which can factor in uncertainty and incomplete information [9]. Material and product selection must be informed by environmental, social, and economic considerations at an early-design stage [10]. The EoL of these products should be planned as early as the briefing and concept design stage. Further research is also needed to (a) embed LCA methods into the early-design process and underpinning workflow enabled by a seamless Building Information Modelling (BIM)-LCA integration; (b) promote comparative approaches that consider multiple environmental, economic, and social indicators to identify the design alternative with the lowest environmental impact; (c) integrate LCA with computational and analytical techniques that can deal with uncertainty, such as machine learning; (d) provide a means of predicting operational carbon emissions during the early-design stages; and (e) explore and promote the acceptance of an LCA philosophy by designers and practitioners, as well as the adoption of the underpinning methods.

During retrofit and construction: Evidence suggests that energy losses and material waste account for 30% and 40% during the construction stage, respectively [11]. Therefore, research into using alternative construction systems (e.g., prefabrication, modular construction, and 3D printing) may provide a means to reduce the environmental impact during the construction phase. It has been noted that a circular economy approach (i.e., design, use, reuse, and recycle) contrasts with the linear value chain model that is used in the building construction industry [12] and is hindered by several structural barriers [13]. This is exemplified by the management of waste in the industry. Consequently, research is needed into approaches that promote decarbonisation and waste elimination in construction, which involves the complex supply chain that gravitates around a construction site. Furthermore, non-energy-related rehabilitation measures tend to be ignored, while the focus remains on improvement to the building envelope, energy system, and energy end-use [14]. Therefore, holistic (i.e., system engineering) retrofitting approaches that are rooted in an LCA philosophy and are informed by decision support systems (including machine learning and optimisation algorithms) should be promoted.

In the operation stage: Current approaches to LCA do not consistently factor in (in the foreground and background inventory systems) life cycle variations in (a) building usage, (b) energy supply (including from renewable sources), and (c) building and environmental regulations, as well as other changes over the building's or district's lifetime. These include: (a) change in the energy mix of a building or district, or upgrading or retrofitting the energy system(s) in place; and (b) time-increase of energy demand during the lifetime of a building due to a wide range of reasons, including changes in occupancy patterns. In this context, the key limitations and challenges that are faced by the current LCA methods and tools include: site-specific considerations [6], several local impacts need to be considered in building assessments (e.g., the microclimate); (b) model complexity [5], buildings use a wide range of material and products, which interact as part of a complex assembly or system; (c) scenario uncertainty [5, 6], the long use phase of buildings (including the potential for future renovation) poses uncertainty problems in LCA that are currently not addressed; (d) health and well-being [6, 7], traditional LCA methodologies do not address the indoor and outdoor environmental impact on health and well-being; and (e) there is a lack of consideration for the social and economic aspects [5, 8].

LCA in EoL: The recycling stage of a built asset is increasingly attracting research, fuelled by the need to promote circularity principles. There is even a growing trend to use the 'cradle-to-grave-to-reincarnation' concept in the recent literature. However, an efficient recycling strategy should be embedded during the early concept design stage of a built asset. Furthermore, it is interesting to note that existing databases (e.g., Ecoinvent) are incompatible with the EoL treatment of the widely-used LCA methods, including Product Environmental Footprint (PEF) and CEN EN 15804/15978 [15]. Therefore, future research should: (a) enhance existing Life Cycle Inventory (LCI) databases to embed EoL data and information; (b) promote comparative approaches that consider multiple environmental, economic, and social indicators to identify the optimal material selection and design alternative with the highest recycling potential; and (c) promote the use of semantics and digital twins of built assets to facilitate the dismantling and reuse of building parts.

1.3 Research Motivation

The current LCA methods have a number of limitations and gaps, including:

- Lack of reasoning and decision-support capabilities, such as exploring "what if" scenarios for the evaluation of alternative design options and devising adapted strategies, thus promoting the active control of buildings and districts [7].
- Lack of alignment with domain models, such as BIM, Geographical Information Systems (GIS), and LCA data structures [16, 17].
- Lack of full support of temporal information [5, 6, 18, 19]. There is also a need to factor in temporal information in the background and foreground LCI and Life Cycle Impact Assessment (LCIA) phases to address maintenance, operation, deconstruction, disposal, and recycling stages.

Recent research has used more advanced approaches to LCA [7, 8], such as incorporating economic considerations by including Life Cycle Costing (LCC). In addition, there is a growing interest in the integration of BIM with environmental impact calculation methods [17]. However, this work is currently limited by semantic incompleteness and interoperability issues between current software solutions. In addition, efforts to scale up LCA from building to district levels are still limited [5, 17].

The application of LCA for buildings requires informed interventions to achieve carbon neutrality, including the elaboration of carbon-intensive activities. These decarbonisation strategies use optimisation approaches to reduce material and energy demand, while integrating renewables and achieving a higher order of efficiency of resources. A carbon neutrality assessment can be also applied and scaled to a district level by adopting reduction and avoidance strategies, and by adapted analysis of the value chain [20].

1.4 Research Objectives

The central aim of this thesis is to develop a semantic-based approach to deliver a near real-time environmental footprint, and inform effective operation and management strategies for the built environment. The hypothesis to be tested states:

A semantic-based approach can facilitate the process of LCA and improve the accuracy of the LCA results by leveraging the value of dynamic data, learning systems, and digital built-environment resources

To evaluate the hypothesis, this research has formulated the following research questions:

RQ1. What are the key limitations of current LCA methods that affect the accuracy and widespread adoption of LCA in the building domain?

RQ2. Can access to dynamic data provide more accurate accounts of the environmental impact during the operation stage?

RQ3. How can machine learning and optimisation be leveraged to reduce the environmental impact of buildings?

RQ4. Can a semantic web approach provide a sound basis to facilitate and streamline the application of LCA in buildings?

1.5 Research Contribution

During the course of the PhD program, the author participated in a collaborative research project entitled “SemanticLCA”. It is pertinent to note that the author held an independent scholarship that supported the PhD studies, and there was no financial compensation from the research project. Within the context of the SemanticLCA project, the author pursued a distinct thread of work that ran in parallel to the engagement in different work packages of the research project. In essence, the project served as a thematic compass and provided access to valuable resources, such as data repositories, and feedback from domain experts and industry partners.

The research trajectory in this thesis was guided by a thorough literature review carried out by the author, which was later published. A key part of this thesis was developing a framework to define the problem and the requirements for delivering a specific use case investigated in Chapters 4, 5, and 6. While the project team gave valuable inputs in refining the work, the framework development and the validation of the proposed solution were the author’s individual contributions.

This thesis makes two main contributions. The first contribution is the development of an integrated framework for optimising building energy and environmental performance, including the development and testing of two machine-learning models and an optimisation strategy for controlling ventilation systems in buildings to improve the accuracy of the LCA results.

The second contribution is the development of a lightweight ontology for LCA that is applied to buildings using a modular approach, which improves interoperability and information exchange across different domains by facilitating semantic modelling.

1.6 Thesis Overview

The thesis is organised into seven chapters (including this introduction). The chapters aim to answer the research questions that were posed earlier. The present chapter has established the context of the research, provided background information, and highlighted its significance.

Chapter 2 will review the literature. It commences by examining the state-of-the-art LCA research applied to buildings, concentrating on current research directions. It then investigates building energy performance from various viewpoints, including the energy performance gap, and the utilisation of machine learning and optimisation to enhance building energy performance. This is followed by a review of semantics in the context of LCA. Finally, this chapter identifies the research gaps and engages in a thorough discussion of these topics, which were considered to address research Question 1.

The methodology that is used in this thesis is presented in Chapter 3. This chapter begins by discussing the theoretical research philosophy that underpins the study. It then gives an overview of the research approaches that have been undertaken to address the research questions.

The framework that was developed to minimise the environmental impacts of buildings is presented in Chapter 4. This chapter provides a detailed explanation of the modelling techniques that are used in the framework, including prediction optimisation, simulation, and LCA models.

The results and outputs generated by the techniques that were presented in the previous chapter are discussed in Chapter 5. This chapter also provides an overview of the assumptions that were made during the development of these techniques and discusses how they were used to address research Questions 2 and 3.

Chapter 6 investigates the role of semantic modelling and interoperability in automating and streamlining the LCA process in the building domain. In particular, this chapter explores the way in which semantics can be used to address research Question 4.

Chapter 7 concludes this thesis. It summarises the main findings and contributions from the perspective of the research questions. In addition, this chapter discusses the research limitations and makes some recommendations for future research.

Chapter 2

Literature Review

This chapter will provide a comprehensive review of the literature on LCA in the building domain. The main objectives of this review are to identify the best practices for the implementation of LCA in buildings, and to highlight the gaps and limitations in the current applications. In particular, this chapter aims to inform the development of an overarching research approach that can address the research questions that are posed in this thesis. However, the scope of the use case that is investigated throughout this thesis extends beyond the application of LCA alone. To fully understand the context through which LCA can be applied, it is necessary to first examine other research areas, such as building energy performance, machine learning, and semantic technologies. Therefore, this chapter will also consider these areas and their potential impact on the development of the overarching research approach. It is important to note that the inclusion of these areas in the literature review is not intended to identify gaps in these areas. Rather, the goal is to gain a comprehensive understanding of the broader research context and inform the development of an approach that can effectively address the research questions that are posed in this thesis.

2.1 Life Cycle Assessment

This section aims to identify evidence and best practices for the implementation of LCA in buildings, focusing on the gaps and limitations in the current applications. A set of relevant LCA concepts will be explored, alongside their relationship with existing practices, ranging from responsible design and modelling techniques to embodied impacts and renovation strategies. The integration of LCA with BIM is also examined to demonstrate the value of dynamic environmental impact assessment, with insights for the development of semantic LCA strategies.

2.1.1 Literature Review Methodology

A systematic review of recent literature has been conducted to identify the current research topics and applications of LCA for buildings (see Figure 2.1). The methodology that is used to conduct the review has three main stages:

- **Stage 1:** Identify recent authoritative research publications using established search engines

  The systematic review process was conducted in January 2021, and relevant documents were retrieved from SCOPUS using the following keywords to provide a broad and comprehensive perspective: (LCA OR "Life Cycle Assessment") AND ("Building" OR "Built Environment" OR "Infrastructure" OR "Urban" OR "District" OR "City" OR "Neighbourhood"). Initially, this combination of keywords returned 6748 documents, including journal articles, conference papers, book chapters, and reports.

- **Stage 2:** Screening and retaining relevant publications

  As shown in Figure 2.2, research in LCA of the built environment is steadily growing, especially in the past 10 years. Given the sheer number of documents that have been published annually and the incremental nature of

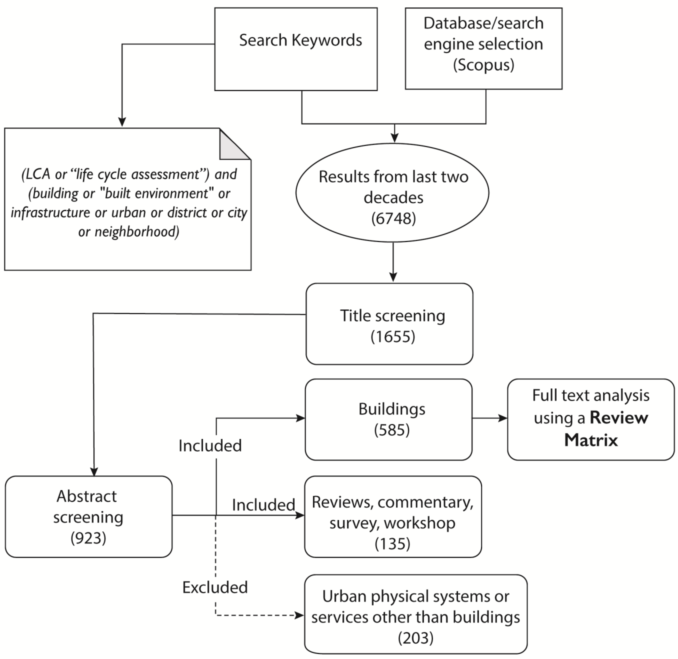


Figure 2.1: Flow chart of the literature review methodology

the published research, this review will only focus on LCA research applied to buildings from relevant recent publications that were published in the last 5 years, while acknowledging seminal work in the past 10 years. The term 'seminal' in this context refers to those articles published in reputable journals and attracted a high level of citations. Also, LCA experts within the project team were consulted to obtain their recommendations for seminal and influential work. Initial screening of the retrieved documents was carried out to identify relevant studies. In this step, the titles and abstracts of 1655 documents were examined to determine whether the study meets the objectives of this review. As a result, a list of 923 documents was created. This list is then divided into three categories: buildings; other urban physical systems (e.g., utilities, transportation system, open spaces, and waste treatment facilities); and existing reviews, commentaries, and surveys (Figure 2.3). Because this study focuses on buildings, studies related to infrastructure and physical assets other than buildings are excluded from further in-depth analysis. Furthermore, studies were included if LCA is directly applied to buildings or to building materials and products.

The chosen approach for this review prioritises recent research on the implementation of LCA in buildings. The decision to focus on recent publications with the last 5 years stems from the understating of the steady growth and evolving nature of research in this field. The substantial number of annual publications make a comprehensive review of the entire body of relevant literature impractical within the scope of this study. Furthermore, it is important to acknowledge that this approach may lead to some blind spots, as it excludes older literature, those lacking the specific keywords used, or publications not indexed by SCOPUS.

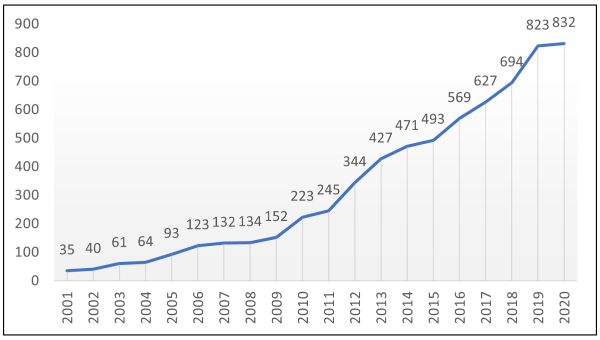


Figure 2.2: Number of built environment LCA publications over the past 20 years.

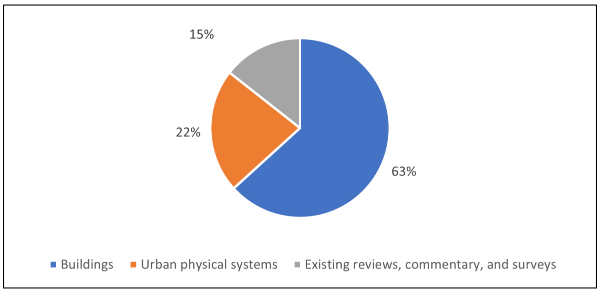


Figure 2.3: Distribution of scientific publications according to scale of application (asset level to urban level) and type of research.

- **Stage 3:** Extracting relevant LCA use cases applied to buildings and analysing their underpinning research.

  A framework has been developed to systematically explore each study to its full extent, which aims to identify the different use cases and highlight the current research trends of LCA for buildings. The following information was collected for each study: (1) scale: this reveals information about building typology and the number of buildings involved in the study; (2) area of application: studies were categorised based on the main objective of conducting LCA (e.g., if a study developed scenarios to enhance the energy performance of existing buildings, then the study is labelled as "energy retrofit"); (3) scope: this gives a brief description of the overall goal of the study; (4) use of BIM and domain models: this aims to identify studies that utilised BIM, or other domain models as part of the framework; (5) utilisation of dynamic data: this aims to capture the use of real-time data in LCA using sensors, smart meters, and IoT devices; (6) consideration of end users or occupants: this identifies the role of human behaviour and feedback on LCA results; (7) impact on human health and well-being: the aim here is to identify studies that have considered the impact of the indoor environment on the occupant's health and well- being; (8) sustainability dimensions: this aims to review the integration of different sustainability aspects (i.e., environmental, economic, and social). The outcomes of the proposed framework will be presented in the following section.

2.1.2 State-of-the-Art Research Landscape in LCA

A thorough review was conducted to elicit the information required by the proposed framework (see Section 2.1.1). Previous reviews on the application of LCA in buildings over the past two decades have identified that most studies focus on energy use and GHG emissions [21–23]. Furthermore, researchers have applied LCA methodology on key areas related to the decarbonisation of buildings. One of the main objectives of the review is to identify the different use cases of LCA applications in buildings. The use cases were identified through an iterative process that extracted the area of application and scope from each identified paper. The second stage factored these findings into a set of generic use cases. Consequently, the structure of this section will follow a use case-based approach. This approach helps to provide an overview of each particular application of LCA, evaluate the current progress, and identify the key challenges and limitations of each area. Figure 2.4 reveals the most common use cases of building LCA. This figure also shows the number of LCA studies per use case. The following subsections will elaborate on each identified use case, starting from the most highly researched.

2.1.3 Environmentally Responsible Design

LCA is increasingly being applied to evaluate the environmental impacts of buildings during the design phase. However, a number of aspects must be considered when performing LCA at the design stage, such as the need for rapid assessment of design variants [24]; the lack of available information, especially in the early-design phase; and the other aspects of sustainability, such economic and social

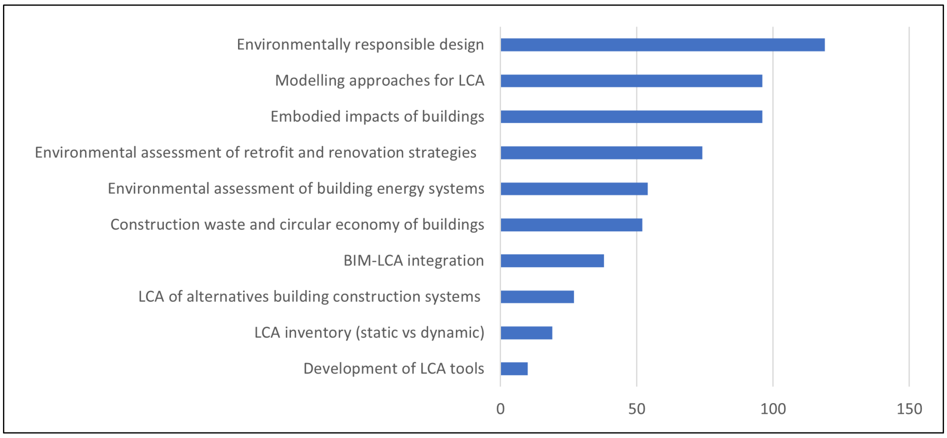


Figure 2.4: The number of recent LCA studies conducted on issues related to buildings.

dimensions. This review has identified three categories of LCA application during the design stage, namely: frameworks, comparative LCA studies, and integrating LCA with other modelling techniques.

The first category includes studies that have developed frameworks to facilitate the workflow of conducting LCA during the design stage, and have proposed a simplified screening approach to select material and structural systems during the early-design stages [24]. The computational workflow assesses the environmental impacts of various configurations of building design and it helps designers to make environmentally-informed decisions, especially when the design requirements and material information are vaguely specified. Zeng et al. [25] integrated design, cost effectiveness, and embodied impacts to facilitate the selection of structural and envelope systems during the early-design stages. Asadi et al. [26] introduced a multi-criteria decision-making model that combines structural resilience with environmental and economic assessment. Hasik et al. [27] developed a framework to estimate the impacts of material use, and energy and water consumption by integrating concepts such as LCA, LCC, energy modelling, and seismic loss analysis.

The second category includes comparative LCA studies that consider multiple environmental, economic, and social indicators to identify the design alternative with the lowest environmental impact, such as the environmental performance of various slab systems [28]; assessment of GHGs emissions, and the energy demand of five structural systems [29]; assessment of Window-to-Wall ratio (WWR), which showed that higher WWR results in higher environmental impacts and economic costs, and led to dissatisfied occupants [30]; the impact of structural design methods on GHGs emissions [31]; and the impact of material selection on carbon emissions during design [32].

The third category includes studies that integrate an LCA methodology with computational and analytical techniques, such as ML, optimization, and Data Envelopment Analysis (DEA). For example, Kiss and Szalay [33] developed a parametric multi-objective optimisation approach to minimise the environmental impacts of different building systems, including envelope, heating, and energy systems. In Manni et al. [34], a parametric multi-objective optimisation model was developed to minimise the embodied carbon and maximise solar irradiation by varying building geometry and orientation. Wang et al. [35] developed a trade-off optimisation-based framework for thermal comfort, LCC and the environmental impacts of the building's envelope. Płoszaj-Mazurek et al. [36] built a parametric machine-learning model to predict carbon footprint using basic design parameters such as wall area, roof area, and height. Finally, Tavana et al. [37] used DEA-based LCA to compare the environmental performance of flooring covering systems.

As noted earlier, conducting a thorough assessment of a given building design is challenging during the early-design stages due to the lack of detailed information and the sheer number of input parameters, which make it difficult to explore trade-off solutions [24]. Nevertheless, the reviewed studies show that developing decision support systems using ML and optimisation methods can be useful in certain aspects of the LCA. ML thrives in data intensive applications (e.g., LCA) because it can be used in optioneering and in the decision-making processes to identify the most informative parameters [38], which reduces the cost and time needed to gather the required data. In contrast, optimisation methods are particularly useful in the design process due to their ability to explore potential improvement options.

2.1.4 Modelling Approaches for LCA

This section will discuss some of the methodological approaches that have attempted to solve issues related to the generic LCA framework, such as the treatment of uncertainty, interpretation of LCA results, and the inclusion of other sustainability dimensions.

2.1.4.1 Interpretation of LCA Results

Reporting and drawing conclusions that are based on quantified environmental metrics that do not always correspond to absolute target values, such as planetary boundaries, is a standard practice in LCA studies. To address this issue, Andersen et al. [39] developed a top-down approach to determine whether or not an environmentally-optimised building design falls within some absolute values, such as the Earth's carrying capacity and the planetary boundaries. Their findings indicate that resource reuse and recycling, as well as reducing operational energy use, are the most effective strategies to meet sustainability goals. Another top-down approach was proposed where the building industry is assigned a share of a country's overall carbon budget [40]. Meanwhile, Rucinska et al. [41] set the target values for the building sector by focusing on local regulatory requirements and the environmental performance of existing buildings to statistically determine the benchmark values. Similarly, Rasmussen et al. [42] calculated reference benchmarks for residential buildings using national samples. The authors emphasised the importance of having consistent calculation rules and a transparent benchmark framework. Another challenge when interpreting LCA results is that environmental indicators are difficult for stakeholders to understand, especially non-LCA experts; hence, the concept of monetary valuation of environmental impacts was introduced. Schneider-Marin and Lang [43] investigated several monetary valuation models and applied them to the embodied impacts of six German office buildings. The authors found that the most important environmental indicators that are recognised by the construction industry are Global Warming Potential (GWP), resource depletion, and acidification potential.

2.1.4.2 End-of-Life Treatment

Enabling the circular economy in the building sector presents the LCA community with methodological challenges regarding EoL treatment, and the allocation of benefits and burdens across multiple-life cycles of products and materials. Eberhardt et al. [44] noted that the existing allocation approaches significantly differ in the distribution of impacts between cycles and their allocation of incentives is questionable. Consequently, they proposed a theoretical model that is based on an existing approach (e.g. linear regressive) to support the transition towards a circular practice. Following a review of two widely-used LCA methods, namely PEF and CEN EN 15804/15978, it was found that the existing databases (e.g. Ecoinvent) are incompatible with the EoL treatment of both methods [15]. The authors also argued that harmonising the two methods is important to obtain more comparative and reliable LCA results.

2.1.4.3 Uncertainty

The difficulty in conducting an environmental assessment of a product is that practitioners often work with incomplete and unreliable information, and in some cases they have to work with unascertained information [9]. This leads to various levels of uncertainty in LCA results. Several studies have attempted to categorise and describe uncertain sources in LCA studies. For example, the International Reference Life Cycle Data (ILCD) Handbook [45] identified three sources of uncertainty: stochastic uncertainty, choice uncertainty, and lack of knowledge of the studied system. Meanwhile Zhang et al. [46] identified three types of uncertainty in the literature: model uncertainty, scenario uncertainty, and parameter uncertainty.

Researchers have addressed the issue of uncertainty using a wide range of approaches. Table 2.1 describes the numerous uncertainty sources in LCA for buildings, the calculation methods applied to quantify uncertainty, the input parameters used in the calculation models, and the extent to which each source of uncertainty contributes to the building's overall impact. For example, Goulouti et al. [47] applied a probabilistic approach to determine the replacement rate of building's elements considering their service life, while Ianchenko et al. [48] used a probabilistic survival model to address the uncertainty of a building's service life. Morales et al. [49] assessed the uncertainties associated with the replacement stage considering the service life of a building's elements and LCI data quality. Meanwhile, Harter et al. [50] studied the impact of a building's development level and shape on the level of uncertainty in Life Cycle Energy Assessment (LCEA) during the early-design stage using a variance-based approach. Resalati et al. [51] examined the effect of embodied energy data uncertainty on the total carbon emissions for the design of a building's envelope. Other researchers have modelled the uncertainty of embodied CO_2 emissions of different building materials considering a building's lifetime and transport distance [52]. In Ylmén et al. [53], a framework was developed to manage choice uncertainty (e.g., design options) in the early-design stages.

Table 2.1: Reviewed uncertainty studies related to building LCA

References	Uncertainty source	Calculation method	Main input parameters	Life-cycle stage contribution
[47]	Replacement rate, reference service life of the building	Probabilistic	Service life of building element	36% of GHGs emissions is attributed to the replacement stage
[50]	Building development level, building shape	Variance-based method	Geometrical technical window, building operation system efficiency	-
[49]	LCI data, service life of building elements	Monte Carlo simulation and scenario-based	Replacement scenarios	The developed scenarios, life-cycle data, and impact categories influence the results of the use stage contribution to the overall impact
[53]	Choice uncertainty	Structured approach and Monte Carlo simulation	Design options	-
[52]	Lifetime of building, transport distance, inventory CO_2 emissions	Probabilistic-Monte Carlo simulation	Building material	-
[48]	Service life	Probabilistic approach	Building lifespan	-

2.1.4.4 Dynamic LCA

A Dynamic Life Cycle Assessment (DLCA) framework has four elemental dynamic components, namely consumption data, basic inventory data sets, characterisation factors, and weighting factors [54]. Using DLCA, Rosse Caldas et al. [55] evaluated the impact of climate change on the environmental performance of a bamboo bio-concrete building considering several factors, including the anticipated increase in temperature, changes in the grid mix, and dynamic characterisation factors. A dynamic weighting system was developed to support time-dependent environmental and planning policies [56]. Meanwhile, Zieger et al. [57] conducted a comparative study between static LCA and DLCA by considering the temporal dynamics of GHGs. It was found that static LCA, combined with other factors, leads to misleading conclusions regarding bio-based materials; however, the DLCA model is more realistic because it considers the timing of GHGs releases and uptakes. Similarly, Negishi et al. [58] noticed significant differences in the results when both static and dynamic models were used, particularly for bio-based materials.

2.1.5 The Embodied Impact of Buildings

Concerns related to the environmental impact of operational energy use in new buildings are now diminishing as a result of effective energy retrofit strategies [59]. However, a major consequence of the enhanced energy efficiency of buildings is the increase in embodied impacts thanks to the additional materials that are required, which transfers the environmental burden from the use phase to the other phases [60]. Therefore, focusing on material efficiency is critical if we wish to mitigate the environmental impacts of buildings [61]. Several material efficiency strategies have been identified, including intense use and lifetime extension of buildings, the use of lighter and low carbon construction materials, minimising construction waste, and the reuse and recycling of building components [10]. As previously mentioned, one of the key aims of current research is to reduce the embodied emissions of construction materials.

Table 2.2 identifies the most common building materials and summarises the main objectives of the reviewed studies. For instance, Kylili and Fokaides [62] investigated the environmental benefits of alternative construction products that incorporate recycled or natural materials. When compared to other building materials, timber has a lower environmental impact and the added benefit of carbon sequestration [63]. Moreover, there is growing interest in alternative bricks that are produced with organic and inorganic wastes that originate in other industries, while research on traditional bricks is decreasing [64]. However, while alternative building materials have many environmental benefits, understanding the extent of their impact is a key barrier to their adoption, together with other important considerations such as reducing costs and eliminating regulatory barriers [65]. Although substantial reduction in GHGs emissions can be achieved from a technological perspective, other aspects of material efficiency strategies must be considered, namely: economic, social, and environmental [10]. Intensive use of building materials and the lifetime extension of products are the most effective material efficiency strategies that have been identified in [10].

Table 2.2: Studies on LCA of common construction materials

References	Material	Objectives
[66–74]	Concrete	Evaluating the environmental impacts of concrete using recycled aggregate and other waste materials, fly ash, steel slag, kaolin clay, and bio-based materials. Comparative LCA of concrete with other materials, such as steel
[75–79]	Wood	Focusing on production of wood-based products (e.g. CLT), logistical challenges and the environmental assessment of timber construction.
[80–88]	Insulation materials	Economic and environmental assessment of insulation materials. Selection of thermal insulation using optimization approaches. Evaluation of bio-based insulation materials.
[89–92]	Phase change materials	Environmental assessment of using PCM for thermal application and energy savings.
[93–98]	Cement	Mostly related to cement production and cement replacement materials.
[99–101]	Earthen materials	Environmental and thermal assessment of alternative building products such as rammed earth and compressed earth blocks.

2.1.6 Environmental Assessment of Retrofit and Renovation Strategies

Measures to rehabilitate the existing building stock are generally applied to enhance thermal performance and reduce operational energy use. In particular, the main focus of building rehabilitation studies is the improvement of the building's envelope, energy system, and energy end-use, while non-energy-related rehabilitation measures are usually ignored [14]. Similarly, Vilches et al. [102] found that energy retrofit, primarily through increased insulation, is the most commonly applied measure, while structural repairs are mostly overlooked. Although the aim of energy retrofitting is to reduce energy consumption during the use phase, the environmental impacts of the applied retrofit measures differ significantly across the life cycle stages [103]. Galimshina et al. [104] applied statistical methods to select the most efficient renovation measure under environmental and economic considerations. Meanwhile, similar retrofit scenarios have been considered and DEA has been used in combination with linear regression to select the most efficient retrofit scenario [105]. Other researchers have utilised Artificial Neural Networks (ANN) to determine the near-optimal energy retrofit scenario by taking into account the environmental impacts, costs, and energy consumption [38]. Rather than considering different retrofit measures, Pittau et al. [106] carried out a comparative LCA of several bio-based insulation materials that have been used on the exterior walls of European housing stock.

Table 2.3 provides more details of studies of LCA-guided building retrofit solutions. Details are provided for each study regarding the retrofit proposals, scale of application (e.g., individual buildings vs district or urban level), the models and analytical methods that have been employed to evaluate the proposed solutions, and the parameters that have been used to estimate and optimise the environmental performance of each retrofit measure. One of the most noticeable differences between small-scale applications (i.e., building level) and large-scale applications (i.e., district level) is the level of data granularity. While studies of individual buildings have been able to utilise more detailed parameters (e.g., heating set point, wall thickness/ characteristics, and operational schedules), studies at the district level have resorted to more generic attributes (e.g., the floor area and the number of stories). Consequently, the accuracy and reliability of LCA outcomes significantly differ. Therefore, new methods are required to provide accurate accounts of the environmental impacts of buildings when considering LCA at the district and wider levels. For example, this may involve the reliance on simulation models that can be developed based on a typology of buildings within a district. This can be facilitated by the use of BIM as well as having access to historical data.

2.1.7 Construction Waste and the Circular Economy of Buildings

The circular economy is a system that seeks to keep materials and products in use for as long as possible, while minimising waste generation [110]. However, closing the energy and material loops through a circular model (i.e., design, use, reuse, and recycle) conflicts with the linear value chain model that is used in the building construction industry [12]. In addition, implementing a circular economy in the building sector is hindered by several barriers, including the fact that the building industry is conservative and fragmented, the lack of a unified and comprehensive framework, and because buildings are usually developed under time and cost constraints [13]. Hence, realising the benefits of a circular economy in the built environment requires changes to be made to the industry practice [111].

Several studies have examined construction waste recycling and component reuse. Ajayebi et al. [112] developed a spatiotemporal mapping model to analyse the po-

Table 2.3: LCA use cases of building retrofit measures

References	Intervention scenarios	Scale	Decision criteria and method	Parameters
[104]	Heating system, exterior wall, windows	Three residential buildings	Statistical analysis of environmental and economic costs	Component types and service life, investment costs, operation costs, user-related parameters.
[105]	Heating system, roof insulation, exterior wall insulation	Residential building	Data envelopment analysis taking into consideration the economic costs and environmental impact	Heating and cooling system set points and efficiency, insulation material conductivity and thickness, exterior wall thickness and conductivity, windows system configuration.
[106]	Exterior wall insulation	Housing stock	Comparative LCA focusing on climate change impacts	Speed of renovation, service life, wall area, thermal performance, type of insulation material, thickness, and so on.
[38]	HVAC system, external wall, roof, façade type, window frame type	University building	ANN taking into account energy consumption, LCC and LCA	Roof surface, exterior wall characteristics, airtightness, operation schedule, temperature setting, space allocation, window design and so on.
[107]	Installation of PV panels, use of renewable energy, minimizing embodied impacts of materials	17 office buildings	Statistical approach, selection is based on environmental impacts only.	Gross floor area, location, roof area, stories, façade materials, window-to-wall ratio, window type.
[108]	Thermal insulation using different materials	672 archetypes of EU residential building stocks	Selection is based on the energy and environmental performance of the studied materials	Location, floor area, number of stories, Story height, window-to-wall ratio, number of occupants, U-value.
[109]	Eight strategies applied to window systems; external and internal thermal improvement; solar thermal system; air chamber insulation; PV panels on roofs; heating system improvement	Residential building	Based on four environmental and economic indicators: non-renewable energy use reductions, net energy ratio, IRR, life-cycle payback	Insulation layer composition and thickness, panel area.

tential for the reuse of building structural products in three urban areas. Their model provides critical information (e.g., product geometries, age, carbon emissions, and weight), which are all necessary for the assessment of future reuse scenarios. Meanwhile, Bertin et al. [113] developed a framework to facilitate future reuse and established a material bank for structural building elements. Their methodology supports the design of a reuse concept and uses the BIM framework to increase the level of details and traceability of the load-bearing elements. Brutting et al. have developed an optimisation method for designing structures from a stock of reclaimed elements [114]. In a comparative LCA study, Minunno et al. [115] concluded that the reuse of building components reduces GHGs emissions by 88% when compared to recycling. However, the viability of the recycling and reuse of construction material (e.g., waste bricks as a replacement to natural aggregates, cement binder, or alkaline activation) is contingent on using advanced technology and rigorous environmental characterisation [116].

In addition to the environmental benefits of implementing circular economy strategies, the economic costs must be considered. Üçer Erduran et al. [117] found that the environmental impacts of new construction using reclaimed wall pieces are lower when compared to the use of new bricks. Meanwhile, the construction costs of using reclaimed bricks are roughly twice the costs of new bricks because the reclaimed wall pieces require the use of expensive equipment. Moreover, the higher costs that are associated with reused elements is attributed to the additional requirements of sampling, testing, design modifications, and the limited supply of second-hand building products [118].

2.1.8 Environmental Assessment of a Building's Energy Systems

Apart from their energy efficiency, sustainable buildings must also produce energy on-site from renewable sources [119]. Previous studies of on-site energy production technologies have provided insights into the costs and benefits of increasing energy self-sufficiency. Table 2.4 provides a summary of the recent studies that have considered the use of renewable energy sources, such as PV systems, energy storage systems, ground source heat pumps, and fuel cells. Table 2.4 also provides information about the use of energy storage technologies, the location of the installed system relative to the building, and the main findings of the study.

Very few studies have considered the impact of operational energy use. González-Prieto et al. [120] found that the operational energy to total impact ratio varies considerably depending on the following three factors: thermal energy source, local climate, and the building's shape. Gardezi and Shafiq [121] developed a linear regression model to predict carbon emissions from operational energy using four variables, namely: construction area, building volume, building lifespan, and weight. Although this study did not comment on the significance of each variable, it does provide an approach for predicting operational carbon emissions during the early-design stages. Meanwhile, other factors that affect the environmental and economic impacts of energy consumption have also been studied. For example, Walzberg et al. [122] considered the possibility of a rebound effect in smart homes because the occupants' energy consumption behaviour is primarily influenced by economic rather than environmental signals. The authors recommended the inclusion of environmental signals in the smart management system because the agent-based simulation model shows a five-fold increase in the rebound effect when load-shifting is driven solely by an economic signal. In addition, O'Rear et al. [123] compared the effect of heating fuel type, specifically natural gas and electricity, on the sustainability performance of buildings. The authors found that electric equipment is more likely to achieve net-zero energy performance, while having higher environmental impacts.

2.1.9 BIM-LCA Integration

BIM can be seen as a resource for information because it virtually represents both the physical and functional characteristics of an asset [136]. BIM is an established enabling technology for the architecture, engineering, and construction industry(AEC) that integrates all of the project phases, and facilitates commu-

Table 2.4: LCA studies of renewable energy systems for buildings

References	**On-site energy production**	**Storage**	**Technology used**	**Location relative to building**	Findings
[124]	✓	-	The application of transparent PV	Window and skylight	The solution reduces energy use
[119]	✓	✓	PV-battery integrated system	-	Less environmental impacts given on-site energy production and components of the battery are recycled or reused.
[125]	✓	✓	PV-battery integrated system	roof	Low-energy design outperforms the net zero energy building approach
[126]	✓	-	PV	Roof	From an environmental perspective, the performance of roof-mounted PV vs commercial PV farm is scenario dependent. Commercial farm performs better that roof-mounted PV panels, but the opposite is true when long distance transmission is required
[127]	✓	-	Ground source heat pump	-	Two types of GSHP: ASGSHP and VGSHP; from environmental point of view, AGSHP causes less environmental impacts.
[128]	✓	-	Heat pump, biomass boiler	-	Both systems have similar impacts, with the biomass boiler causing more impact during manufacturing
[129]	✓	-	Dual source heat pump	-	Most of the impacts occur during the use phase (i.e., electric energy consumption).
[130]	✓	-	PV	Roof	The environmental benefits of a rooftop PV system are dependent on local electricity mix and installation location; payback period is estimated at 11 years.
[131]	✓	✓	PV-battery integrated system	Roof	This study developed an optimisation model for optimal PV-battery sizing.
[132]	✓	✓	Stand-alone power plant	-	The environmental impacts of manufacturing and use stage are driven by the location of the building.
[133]	-	✓	Aquifer thermal energy storage and in-situ bioremediation	-	The environmental impacts of the proposed system are less by a factor of two compared to the conventional heating and cooling system.
[134]	✓	✓	PV-Battery integrated system	-	The study focused on the environmental and economic optimal configuration of solar systems.
[135]	✓	✓	PV-Battery integrated system	-	This study compared different battery technologies for residential applications.

nication and information exchange among project teams [137].

BIM is viewed as a technology that facilitates LCA application in the building sector by providing integrated solutions to a data-intensive and time-consuming method, such as the LCA [138]. Figure 2.5 shows that there has been a marked increase in the number of LCA studies that have utilised BIM using different workflows, especially over the past 5 years. The integration of BIM and LCA can automate the exchange of information and facilitate data acquisition between BIM models and LCA databases and tools, which significantly accelerates the environmental assessment of design alternatives during various stages of the building's life cycle [17, 139].

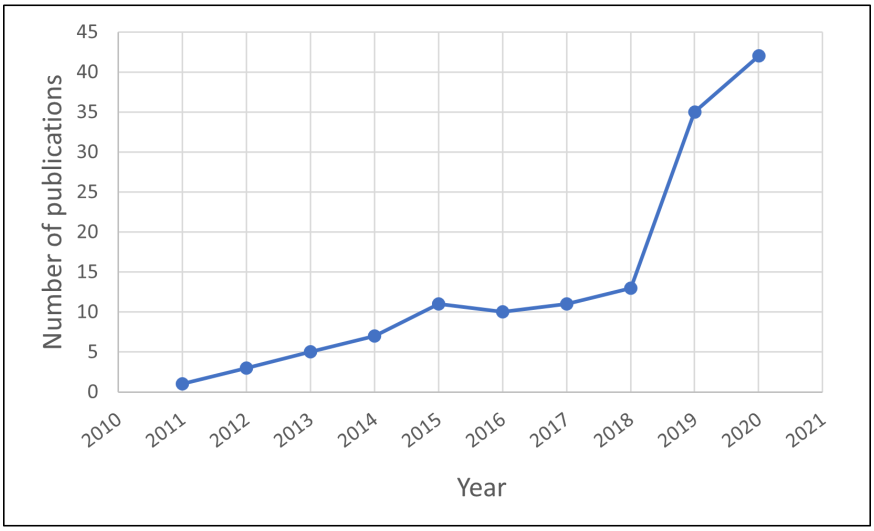


Figure 2.5: Number of BIM-based LCA studies from 2011 to 2020

2.1.9.1 BIM-LCA Integration Approaches

Several classification schemes for existing BIM-LCA integration workflows have been proposed. Each proposed classification scheme uses different features and characteristics to identify the existing integration approaches. Soust-Verdaguer et al. [17] identified three levels of integration on the basis of data input, data analysis, and communication of results. In the first level, BIM is only used to quantify and identify the building materials (i.e., quantity takeoff) during the LCI step. In the second level, environmental data are embedded in the BIM model. Finally, the highest level integration is achieved by creating an automated workflow that combines various data sources and software. Wastiels and Decuypere [140] proposed a more comprehensive classification by considering the integration workflow and the direction of the data flow. They identified five main strategies, as follows: i) extracting bill of quantities (BOQ) from the BIM model; ii) importing the geometric BIM model to a dedicated LCA tool; iii) using a BIM viewer to attribute LCA profiles to the BIM objects; iv) using a plug-in to perform LCA within the BIM environment ; and v) embedding LCA data in the BIM model. Nizam et al. [141] categorised BIM-based LCA studies into four types based on the clarity and applicability of the proposed framework, the role of BIM, and the scope of the LCA calculation (e.g., exclusion of some life cycle stages). The most recent review by Safari and AzariJafari [142] used the number of required manual inputs and the level of complexity of the integration process (e.g., data exchange and computation types) to classify existing approaches into three categories, namely: conventional, static, and dynamic.

In this thesis, the reviewed studies are classified into three major categories, namely: using BIM to extract the material quantities (BOQ) (type I), integrating BIM with the LCA methodology via plug-in tools (type II), and embedding environmental data in the BIM objects (type III). Furthermore, as shown in Figure 2.6, another layer of assessment was considered to evaluate the integration approaches. In a true integration, a permanent bidirectional link is established between the BIM model and the LCA calculation. Consequently, the proposed framework can incorporate future design iterations with minimal additional effort by the users. Meanwhile, in a loose integration the entire workflow requires an intensive manual reworking to incorporate changes to either the building design or the LCA parameters. As seen in (Figure 2.6), frameworks of type I are con-

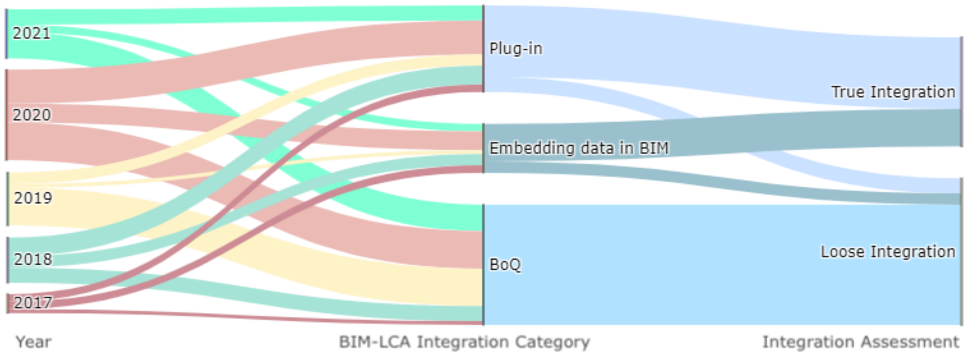


Figure 2.6: Visualisation of the integration approaches of the reviewed studies

sidered as loose integration, while the type II and type III frameworks are seen as true integrations. A more detailed discussion of each integration approach is provided below.

BOQ

Several studies have used the BIM model to extract essential building data, such as the material types and quantities, and building geometry (i.e. BOQ) for conducting LCA. One of the most recognised approaches for integrating BIM with environmental performance assessment is exporting BOQ from BIM [17, 143]. In this approach, BIM is mainly employed to establish LCIs by calculating material quantities and then exporting them as spreadsheets. Hence, this approach is considered to be a loose integration due to the excessive manual effort that is required to manage and map the data, and the iterative process that is required to account for the changes during various design stages. Su et al. [144] proposed a method to evaluate the environmental impacts of demolition waste using BIM for automatic extraction of the building materials. In [145], a BIM-based framework was developed to identify the trade-off between building operation and embodied impacts by exporting geometric data to the Athena LCA calculator. Another study established LCI by integrating quantity take-off from a BIM model with a subset of ÖKOBAUDAT, which is a German LCA database for buildings [146]. Other researchers have integrated multiple data sources (e.g., material quantities from BIM, EPDs, and construction operations) to evaluate the embodied impacts of various design options from cradle-to-grave, and of potential recycling and reuse scenarios [147]. Carvalho et al. [148] advanced this approach and developed a Dynamo routine to calculate LCA by establishing a link between the bill of quantities and the materials' life cycle impacts. Both data sources were stored as Excel spreadsheets and the linking mechanism between the elements in the two data sets is based on the same material name.

Overall, this approach has been criticised as being static in nature and it typically creates a one-directional workflow that results in non-interoperable systems [141, 149]. Furthermore, as noted in [143, 150], this approach is inadequate for conducting a whole-building LCA, or to compare multiple design options because the process of transferring and mapping data between different tools is complicated, time-consuming, and error-prone.

Plug-in Approach

BIM-integrated LCA tools, mainly in the form of plug-ins, have been widely used to integrate BIM with LCA methodology. These plug-ins can be proprietary, such as *Tally* LCA, or they can be experimental developed for research purposes. Tushar et al. [151] developed an integration between *Autodesk Revit*, *Tally*, and an energy rating tool to reduce the environmental footprint and energy consumption of various design options. Another study used the *Tally* plug-in to conduct an environmental impact assessment of prefabricated concrete components [152]. In [153], a framework was developed to enhance the selection of building envelopes by integrating BIM, LCA plug-in (*Tally*), and an optimisation model. In [139], a combination of BIM, an LCA tool (*Tally*), and machine-learning algorithms was applied to evaluate several building typologies and to identify the key design variables that influence the environmental performance of buildings.

Sameer and Bringezu [150] integrated the data exchange between the BIM environment (*Autodesk Revit*) and the footprint data of building materials during the design stage using several APIs. In the proposed framework, the openLCA API was used to establish the environmental footprint of construction materials and another API was developed within *Revit* to facilitate the data exchange, and quantify and visualise the environmental footprints within BIM. Nizam et al. [141] developed a Revit plug-in to estimate the embodied energy of materials, transportation, and construction within the BIM environment. This framework connects information from the BIM model to a designated database that contains the embodied energy coefficient. Similarly, a Revit plug-in has been developed to enable data mapping between the extracted materials from the BIM model to a customised subset of the Ecoinvent database [154]. After establishing the LCI, openLCA software was used to undertake the LCA. Rather than using external LCA software, Kiamili et al. [155] used Dynamo to establish a bidirectional link between BIM objects and a customised LCA database. Their proposed workflow links the extracted building materials to their corresponding LCA data using material-based mapping and incorporates design changes for real-time environmental assessment within the BIM environment. Similarly, Hollberg et al. [156] used the material IDs that are defined in the KBOB database to link building materials from the BIM model with the associated LCA data in the KBOB. When the KBOB IDs are added to the BIM object, a Dynamo plug-in multiplies the BOQ by the embodied factors from the KBOB database.

Researchers have raised some concerns about the existing BIM-compatible LCA tools because they exhibit a compromise between simplicity and transparency. The LCA results that are generated by these tools can be viewed as a black box [157], in which the end users have minimal knowledge about the internal workings and assumptions. This can prevent a deeper understanding of the LCA results and a thorough understanding of environmental hot spots [158]. Another study has found inconsistency in the results of a dedicated LCA tool (GaBi 6) and a BIM-LCA plug-in [157]. The authors hypothesised that the discrepancy between the two methodologies is due to the simplifications of the plug-in tools that are used to allow non-LCA experts to conduct an environmental assessment during the design stage. For the same design problem, various plug-ins can produce incomparable results because each plug-in tool uses different databases and workflow. For instance, *Tally* makes use of the GaBi database and allows for LCA calculation at a different level of aggregation (i.e., a whole building component vs individual layers), while *One Click LCA* provides much wider choices of EPDs and evaluates each material separately [138]. In addition, these integrated tools do not allow users to add specific data records from external sources [159]. Furthermore, although this approach provides simultaneous feedback, the accuracy of the results is questionable because plug-in tools frequently rely on generic environmental data [160]. Nizam et al. [141] identified further limitations specific to *Tally* including the omission of the construction processes, and the lack of automatic data mapping between BIM objects and LCA data.

Embedding Data in the BIM Objects

Instead of exporting BIM data to external LCA tools, this approach suggests embedding all of the LCA-related parameters and environmental factors in the BIM model. This has the advantage of facilitating information reuse and exchange among project stakeholders within a single model and allows for direct reporting on the environmental implications of various design options [161, 162]. In general, two types of information are embedded in the BIM model under this approach: the first type includes the basic parameters that are required to conduct the LCA, such as nature of the resource, energy source, type of transport, recyclability, and so on [163]; the second type focuses on inserting new environmental parameters (e.g., climate change, resource depletion, acidification, etc.) into the BIM model [164]. Establishing the links between multiple data sources requires extensive LCA knowledge; hence, the inclusion of such information within the

BIM environment can potentially benefit non-LCA experts [165].

Ansah et al. created a functional database that comprises a list of building materials and components, as well as the properties required for LCA calculation [166]. To conduct the environmental assessment within the BIM environment, this framework employed Dynamo and Structured Query Language (SQL) to map and insert LCA parameters into the BIM model. In [167], a different strategy was proposed to evaluate the environmental impacts within a BIM authoring tool—the authors created a BIM library of major construction materials that includes environmental impact parameters that are derived from an LCI database. However, the limitation of this framework is that the list of environmental indicators is predetermined and only applies to building materials that were previously selected. Santos et al. [161] demonstrated that by augmenting the BIM model with semantic information acquired from project documents and LCA generic databases, it is possible to streamline environmental and economic analysis throughout the design stage. The authors used the Revit API and information delivery manual (IDM) to manage and handle the information flows of the proposed framework. Horn et al. [162] noticed a lack of transparency and standardisation regarding information exchange requirements of the previous studies using this integration approach. To resolve this issue, the authors created a detailed data requirement structure based on the IFC standard. The proposed solution creates a bidirectional information flow between the BIM software and a dedicated LCA tool. Consequently, the basic LCA input and the LCA results that are generated by the LCA tool will be inserted into the BIM model based on the IDM standard.

Although this strategy incorporates environmental information into the BIM objects to simplify and automate the LCA process, the designers require special training to properly interpret the provided information [159]. Furthermore, embedding LCA data into the BIM model could create heavy and inoperable models [155]. Another issue with the majority of the suggested solutions is that they lack integration with specialised LCA software, which limits the scope of LCA study to a predetermined set of life cycle stages and environmental indicators. Therefore, any modifications to the LCA design parameters (e.g. functional unit, and system boundary) or the addition of new design features and building materials will not be considered. Finally, Nizam et al. [141] argued that the complexity and inadequate details of the existing studies that follow this type of integration make the adoption impractical.

2.1.10 LCA of Alternative Building Construction Systems

The construction industry is known for its energy intensity and high carbon emissions [168]. During the construction stage, energy losses and material wastes are about 30% and 40%, respectively [11]. Alternative construction systems (e.g., prefabrication, modular construction, and 3D printing) can help to reduce the environmental impact of buildings in the pre-use stage. Table 2.5 describes the construction system that have been evaluated, the building materials that have been used, the dimensions of the sustainability being considered, and the main findings of the study.

The use of prefabricated building components lowers carbon emissions and reduces environmental impacts when compared to the cast-in-place method [169–172]. Yao et al. [172] applied a monetisation approach to facilitate a comparison between the environmental and social factors. The authors found that the assembly stage has the highest environmental impact, and that the key contributors are energy and fuel consumption, noise pollution, and the loss of components and materials. The environmental benefits of using a prefabricated building envelope have also been considered [173, 174]. The performance of a modular building envelope depends on the material selection, module design, and the availability of the products within an acceptable distance to minimise the impact of transportation [173]. Furthermore, the production of prefabricated concrete elements (PCE) with recycled construction and demolition wastes lowers GHGs emissions and costs when compared to PCE produced with virgin material [174].

2.1.11 LCA Data (Static vs Dynamic)

The development of the LCI is a core design parameter of LCA methodology, which refers to the collection of data related to the inputs and outputs of a particular product system. There are two main categories of data: primary data, which LCA practitioners collect themselves; and secondary data, where the data are drawn from generic databases or literature. Silva et al. [181] found that the limited adoption of LCA is due to the amount of data that is needed to establish the LCI. The authors proposed that primary data collection should be prioritised to foreground processes because they account for most of the environmental burdens of construction products, while background processes can depend on the existing databases. However, the environmental impacts of building materials and products are quantified using precalculated coefficients from existing databases, which are frequently criticised for being inconsistent and incomplete [182]. Instead, Crawford et al. [182] proposed a hybrid method that combines data from the process-based approach with economic input-output data, which will generate a more comprehensive and accurate LCI.

To provide an accurate environmental assessment, it is vital to build regionalised databases that reflect real-world scenarios. In this regard, Alzard et al. [183] claimed that creating a representative LCI data set for the production of recycled concrete aggregates in a UAE city enabled stakeholders to make informed decisions about whether recycled aggregates are a more environmentally-friendly option. Ayagapin and Praene [184] showed that environmental costs significantly differ depending on a number of factors, such as sources of construction materials, transportation method, electricity mix, and geographical location. This

Table 2.5: LCA application for evaluating alternative construction systems

References	Construction system	Building material	Environmental LCA	Economic considerations	Outcomes
[175]	Prefabricated modules	Steel vs. concrete	✓	✓	Overall, steel shows better performance in terms of environmental and economic factors.
[173]	Prefabricated envelope	Composite system	✓	-	Environmental performance of modular construction depends on material selection, module design, and the availability of the products within an acceptable distance to minimise the impact of transportation.
[170]	Prefabricated building components (e.g., stairs, wall, beams)	Steel and concrete	✓	-	A 15% reduction in carbon emissions compared to the conventional cast-in-place method.
[176]	Prefabricated frames	Steel	✓	-	An anticipated reduction in carbon emissions and energy use by 4.4% and 9.2%, respectively.
[177]	Prefabricated slab	Concrete	✓	-	Significant reduction in carbon emissions by nearly 35% compared to the cast-in-situ method.
[11]	prefabricated concrete piles	Concrete	✓	✓	A linear relationship was found between the construction stage carbon emissions and the area, number, and cost of pile foundations.
[171]	Prefabricated concrete structures	Concrete	✓	-	Overall, the carbon footprint of prefabricated structures is lower than that of cast-in-place structures.
[178]	Prefabricated concrete deep foundations	concrete	✓	✓	The use of prefabricated deep foundations results in a reduction in most impact categories but the prefabrication method incurred higher construction cost (12-37% increase).
[172]	Prefabricated building components	Concrete	✓	✓	The assembly stage has the highest environmental impact. The key contributors are energy and fuel consumption, noise pollution, and the loss of components and materials.
[174]	Prefabricated façade	Concrete	✓	✓	The production of prefabricated concrete elements (PCEs) with recycled construction and demolition wastes has lower GHG emissions and cost compared to PCE produced with virgin material.
[179]	Digitally fabricated building elements	Mainly wood and concrete	✓	-	Digital fabrication techniques can provide environmental benefits and material efficiency during production; however, the use of hybrid materials in multi-functional architectural elements could negatively impact material recyclability.
[169]	Prefabricated building components (stair, slabs, beams, etc.)	Concrete	✓	-	Prefabricated buildings have less environmental impact whyen compared to cast-in-place buildings, by nearly 18%.
[180]	Modular construction	wood	✓	-	Modular construction is not always the optimal choice and practitioners must consider design optimisation, material waste reduction, and transportation needs to improve the viability of this choice.

indicates the importance of regionalised databases. Moreover, Environmental Product Declarations (EPDs) have emerged as a major tool in environmental assessment policies in developed countries, driven by the widespread adoption by several environmental certification systems, regulatory requirements, and as EPDs are increasingly being considered by environmental assessment tools [185].

Data granularity also affects LCA results because the results of some impact categories are strongly related to data resolutions [186]. In addition, granular data can generate more accurate LCA results [187]. The use of real-time data is crucial to the accuracy and reliability of the LCA results because of the dynamic nature of buildings. Vuarnoz et al. [188] demonstrated how the real-time data of occupancy profiles and appliance usage patterns can be used to improve the accuracy of LCA results. Examples of common real-time data sources include smart utility meters [188], Internet of Things (IoT) for occupancy detection and appliance use [189, 190], and sensors to measure indoor temperature and relative humidity.

2.1.12 Development of LCA Tools

Several proprietary and open source LCA tools have been developed to support the application of LCA (e.g., OpenLCA, SimaPro, and GaBi). However, the limited adoption of LCA for buildings can be attributed to the complexity of buildings and the amount of data that is required to establish LCIs [181]. Therefore, various specialised LCA tools have been developed to facilitate LCA practice in the building sector. LCA tools can be built as stand-alone software (e.g., the Athena Impact Estimator) or as a plug-in (e.g., Tally and One Click LCA). Although existing tools can simplify LCA calculation, they are viewed as a black box since the end users have little knowledge about the assumptions and internal workings of the tool [157]. This can preclude a comprehensive understanding of the results of the LCA [158]. Furthermore, different LCA tools can generate inconsistent results for the same design problem because each tool utilises various workflows and databases [138] and may omit some processes, such as construction [141]. For a more detailed discussion of the limitations and challenges of LCA tools for buildings, the interested reader is referred to the following recent reviews [138, 143].

In addition to the existing commercial LCA tools, researchers have also developed solutions to address specific aspects related to building LCA. Domjan et al. [191] developed an Excel-based LCA tool to evaluate operational energy use and embodied emissions. Miyamoto et al. [192] developed a decision support tool to integrate LCA and life cycle cost during the early-design stage for dwellings. To reduce the time required to compare design alternatives, Duprez et al. [193] created Machine-learning (ML) models that allow designers to rapidly evaluate new alternatives using the trained models.

2.2 Integrating Machine-Learning with LCA

There is a growing trend in the literature to use artificial intelligence, including ML, for various LCA applications. ML approaches are capable of enhancing their prediction accuracy without requiring reprogramming by learning from obtained data. This involves the development of a model that can discern patterns from training data and then create an algorithm without the need for human input, as described by Mitchell [194]. Considering the data-intensive nature of LCA, which involves vast numbers of input parameters and associated uncertainties, as well as the costs of data collection and the typically large number of alternatives involved, ML may be a useful tool for supporting LCA. According to a recent review by Ali et al. [195], ML has the potential to be a valuable tool in certain aspects of LCA. Their findings indicate that ML methods can be effectively utilised in optimising scenarios within LCA. Moreover, the integration of ML methods into existing inventory databases was found to streamline the LCA process across various use cases. A systematic literature review that was conducted by Barros and Ruschel explored scientific research in the context of LCA and ML within the architecture, engineering, and construction industries [196]. Their findings reveal that energy consumption and Global Warming Potential were the most frequently investigated environmental indicators. Additionally, the authors discovered that ML was mainly employed for prediction purposes. Consequently, this section will provide an overview of the ML techniques that have been employed in LCA studies, with a particular focus on the building domain.

ANN are a popular ML technique that is widely used in many fields, including LCA. In their study, Shi and Xu [197] proposed a systematic method for performing LCA to evaluate the environmental impact of construction materials. The authors also introduced two ML algorithms—the Back Propagation Neural Network (BPNN) and the hybrid Genetic Algorithm-Back Propagation (GA-BP)—to assess the environmental performance of the studied materials. The results revealed that the GA-BP algorithm outperformed the BPNN in terms of precision and selecting materials with lower environmental impacts. D'Amico et al. [198] employed ANN to address both the energy and environmental aspects of a building's LCA. The researchers created a decision support tool that enables fast and accurate evaluation of a building's performance. The authors showed that the ANN algorithm is useful in predicting both energy demand and environmental impacts in LCA of buildings. Sharif and Hammad [199] devised a surrogate ANN to aid in the selection of optimal building energy renovation methods. The developed ML models were used to generate renovation scenarios that take into account the total energy consumption. Azari et al. [200] determined the optimal design for building envelopes using a multi-objective optimisation algorithm. Their study focused on the energy consumption and life-cycle environmental impacts of an office building. The input variables that they used for the design included window type, window frame material, insulation material, wall thermal resistance, among others. The optimal combination of these variables was obtained to design a building with the least possible environmental impact and operational energy. The authors used eQuest 3.65 to calculate active energy, while Athena IE methods were employed to estimate the LCA. Moreover, an ANN model in combination with genetic algorithms was employed to generate additional design combinations and to identify the ideal design iteration.

Random Forest (RF), another ML technique, has been extensively employed in LCA studies due to its good prediction performance, in addition to its built-in variable importance tool [201, 202]. Frömelt employed three ML algorithms (i.e., RF, k-nearest neighbors (KNN), and LASSO-regression) to attribute missing information on water supply, electricity, and heating [203]. Subsequently, the predicted data were converted into quantities utilising price data. Based on the household budget survey, similar socioeconomic household archetypes were identified in consumption. The divergence of these archetypes from general macro-trends indicates that the proposed approach has the potential to enhance our understanding of consumption patterns, and thereby aid policymakers in making informed decisions regarding impactful environmental policies that target certain groups of consumers. De Rousseau [204] studied concrete mixture design optimisation and compared different ML methods such as regression models. They determined that RF was the most effective technique for predicting the compressive strength of concrete in actual field mixtures. This conclusion was based on an evaluation of the model's performance metrics. The results were used to inform the subsequent LCA calculations. Gu [205] created an LCA model that aimed to minimise the environmental impacts that are associated with metal-organic frameworks. To achieve this goal, the author combined a conventional LCA with RF and obtained preliminary guidelines for sustainable metal-organic framework design.

Other ML techniques beyond ANN and RF have been sparingly employed in the field of LCA [195]. For instance, Support Vector Machines (SVMs) and hybrid ensemble methods, such as the gradient-boosted classifier tree ensemble model (GBM), have not been extensively used. Overall, ML is a valuable tool for several aspects of LCA. However, the accuracy and effectiveness of ML solutions used in LCA heavily rely on the quality and reliability of the underlying database, which remains one of the most challenging aspects of LCA. Hence, the integration of reliable data and ML techniques will significantly enhance the speed and accuracy of LCA applications [196, 198]. In addition, ML can be effectively combined with traditional optimisation methods to improve their ability to rapidly explore and evaluate alternatives.

2.3 Semantic Web Technologies to Facilitate LCA

LCA is an interdisciplinary subject that requires the integration of knowledge from various fields and the utilisation of heterogeneous data sources, which creates barriers for information sharing and reuse [206]. Hence, there is a need for a new approach that will enhance and facilitate data interoperability, which can be achieved through the use of semantic modelling and ontology. Therefore, this section will provide a brief overview of the role of semantic web technologies, including ontology, in LCA, by reviewing relevant studies in this area.

2.3.1 Overview of the Semantic Web

The Internet has had a tremendous impact on modern life since its inception about three decades ago [207]. The Internet was initially focused on serving information but the advent of user-generated content, such as Wikipedia, shifted its focus to becoming instead a platform for connectivity and sharing ideas, which was referred to as Web 2.0. [208]. However, this concept has been replaced in research by the introduction of Semantic Web Technologies, which were proposed by Sir Tim Berners-Lee and the World Wide Web Consortium(W3C) [209]. The primary aim of these technologies is to provide a better definition of the meaning of information on the web.

The fundamental idea behind The Semantic Web and Semantic Web technologies is the utilisation of semantic metadata to revolutionise information and process management [210]. The Semantic Web enables machines to comprehend the context and meaning of content by adding abstraction layers. This can be achieved through the use of a number of semantic web technologies, which mainly include:

- **Resource Description Framework (RDF)**: "RDF is a structure for describing and interchanging metadata on the Web" [211]. Essentially, Semantic Web information is represented in a graph format that consists of nodes connected by edges. Nodes store information, while edges represent the relationships between information stored in the nodes. This form of representation is often referred to as RDF triple. An RDF triple consists of subject, predict, and object. To illustrate this concept, suppose that an article is written by an author. In RDF notation, the article would be represented as the subject, and "written by" would be the predicate that expresses the relationship between the subject and the object, which in this case is the author.

- **Web Ontology Language (OWL)**: "OWL is a language for defining and instantiating Web ontologies" [212]. OWL was introduced to provide complete support for ontology creation, building on the RDF functionality [212]. The primary objective of OWL is to provide support for applications that require processing and use the content of information [213].

- **SPARQL Protocol and RDF Query Language (SPARQL)**: SPARQL allows for RDF data to be queried and manipulated in a manner similar to how SQL permits querying and manipulating relational database [214]. In essence, SPARQL functions as a graph-matching query language by utilising a pattern to match against a given data source. The resulting values are then processed to provide an answer [215].

2.3.2 A Note on Ontology

The use of ontology is central to Semantic Web applications, where an ontology is commonly defined as "an explicit and formal specification of a conceptualisation of a domain of interest" [210]. Conceptualisation refers to the role of ontology in structuring concepts to reflect a world view [216]. The term ‘formal’ means that computer-based reasoning is permitted through the use of some syntax, while ’domain’ refers to a specific set of concepts that are identified using requirement engineering [210].

Ontologies are designed to model the knowledge and concepts that are related to a particular domain. They are not intended to structure data for a particular application but rather to capture the essential concepts and relationships of a domain of interest. An ontology that has been implemented is usually integrated into the backend of a system, serving as a repository for data that captures context, standardises terms, enables rule application, and generates novel insights. In addition to their capacity to capture meaning and to provide standardised terminology, ontologies also enable the deduction of knowledge through explicit statements about a domain [217, 218]. The deduction capability is carried out by a ’reasoner’, which allows the inference of the truth of a statement to be made based on other explicit statements within the ontology. Although comprehensive discussion of ontology is not feasible within this section, interested readers are referred to authoritative sources such as [212, 219, 220] for a detailed understanding of this subject.

2.3.3 The use of Semantics and Ontology in LCA

Several attempts have been made to address various aspects of LCA by leveraging Semantic Web technologies and ontology. For example, an ontology has been developed to represent key aspects of LCA, such as flows, and activities, as well as their properties [221]. The motivation for using ontology is to address the limitations of existing LCA data formats, which only address syntactic interoperability and ignore semantics, leading to inefficiencies in information management and the reproduction of published studies. An ontology-based model has been developed to represent the life cycle of functional products—including processes, flows, and their semantic relationships— which are encoded as an RDF graph [222]. The proposed framework was implemented and tested on the life cycle of a ball bearing, demonstrating its validity and practicality for LCA-oriented ontology-based modelling. The authors acknowledge that their research is preliminary and they suggest that future work should include more LCI analysis, and allocation and recycling considerations. Furthermore, LCA faces challenges when accounting for the spatiotemporal dynamics of LCA activities, which can affect environmental impact estimation. To address this issue, an ontology for modelling spatiotemporal scopes has been developed to enhance interoperability between diverse data sets and enable LCA practitioners to address the impact of spatiotemporal scopes on LCA results [223].

In [206], a novel approach for data providers in LCA was introduced, where a catalogue interface is presented to users instead of a standalone database. A catalogue is used to represent the meaning or semantics of a data resource. Then, a semantic software system is used to interpret user queries and route them to data providers that can provide answers. By using catalogue interfaces, private data can be made more easily discoverable and interpretable for users.

A few studies have focused on developing and utilising ontology-based models to facilitate and improve LCI modelling. Bertin et al. [224, 225] proposed an ontology-driven approach to model LCI and regroup processes into semantic groups using an ontology to store keywords that describe each process. The authors also developed a web application that utilises their approach, which consists of a data management back-end and a front-end for visualising process dependencies in a graph and searching for processes. Their approach aims to offer a more comprehensible LCI database model and a new way to express process dependencies. This approach was demonstrated using LCI data for electricity production in the United States, and was implemented using relational algebra and SQL. However, the authors plan to use OWL and semantic reasoner in the future, and will study the impact on big energy data sets to improve the performance of the proposed approach. Similarly, a methodology has been proposed for automated LCI modelling of chemical manufacturing using ontologies [226]. In particular, the authors presented two ontologies, Lineage and Process, to manage data describing the synthesis pathway and unit processes associated with chemical manufacturing. The ontologies are coupled to facilitate automated inventory modelling for a chemical of interest. The authors then illustrated the proposed methodology with a case study of the production of nylon-6.

It should be noted, however, that these studies are still in the early stages and are limited to specific applications within LCA. While they have demonstrated promising results, there is a need for further research in other areas, particularly in the building domain, where the use of ontologies for LCA has not yet been extensively explored. Another limitation of these studies is the lack of standardisation in the resulting ontologies. The defined concepts and relationships are often suggested on an ad hoc basis and may not be widely accepted or recognised within the field. Additionally, many of the ontologies that were developed in these studies are not published or made available in a format that can be easily accessed or integrated with other ontologies, which limits their reusability and interoperability. In light of these limitations, future research in LCA and ontology should prioritise the development of standardised ontologies that can be widely adopted and applied across different domains.

2.4 Digital Twin and Data Intertwining

In recent years, the digitalisation of the built environment has made significant strides, leveraging technologies such as semantics, dynamic data, and ML. A notable advancement in this domain is the emergence of Digital Twin for buildings.

Researchers have suggested several definitions, frameworks, and methodologies for the development and evaluation of Digital Twin applications. In a systematic review on Digital Twin applications in various domains such as manufacturing, healthcare, aviation, and medicine, Barricelli et al. identified around 30 definitions of Digital Twin [227]. The authors recognised the following semantic categories describing Digital Twin: integrated system; clone; links; description/information; simulation/prediction; and replica/virtual. However, according to Tao et al. [228], the most recognised definition of Digital Twin is "Digital Twin is an integrated multi-physics, multi-scale, probabilistic simulation of a complex product and uses the best available physical models, sensor updates, etc., to mirror the life of its corresponding twin."

Beyond the mere definition of the Digital Twin concept, delving into the underlying technologies, elemental components, and examining their interactions is more important than striving for a singular unified definition. Kritzinger et al. [229] proposed a classification of digital twins based on the level of integration and data exchange, illustrated in Figure 2.7. The first category is the digital model, representing a digital replica of a physical asset that operates independently without any exchange of data between the physical and digital models. The second category, digital shadow, involves a unidirectional flow of data from the physical asset to the digital model through sensors and connected devices. Finally, the third category, digital twin, entails a bidirectional flow of data, establishing a seamless connection between the physical and digital twins through the utilisation of sensors and actuators.

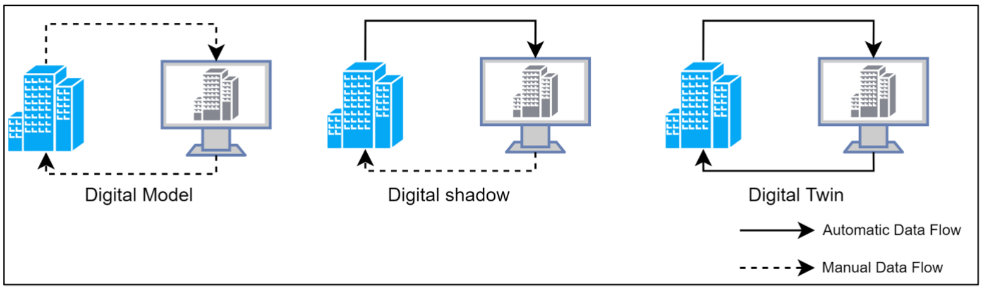


Figure 2.7: Digital Twin categories based on [229]

While the previous categorisation primarily focuses on the level of integration between physical and digital twins, an alternative framework proposed by Qi et al. [230] introduces key dimensions to assess Digital Twin applications. These dimensions encompass the physical entity, a virtual model accurately replicating the actual physical product, real-time data, and services that facilitate simulation, optimisation, and predictive analytics. Enders and Hoßbach [231] introduced a schema for comprehending and constructing Digital Twin applications, which encompasses several essential dimensions: i) creation time of Digital Twin, which provides background information into when the model was initially generated; ii) connection between physical and virtual twins, which evaluates the nature of linkage between the physical and its virtual counterpart, ranging from none to unidirectional or bidirectional connections; iii) physical reference object, this dimension encompasses a diverse range of objects such as assets, products, humans, and infrastructure; iv) purpose, which outlines the primary objectives of the Digital Twin such as simulation, monitoring, or control; and v) completeness, which evaluates the level of details incorporated within the digital model pertaining to its representation of the physical object.

According to Calin et al. [232], the Architecture, Engineering, and Construction (AEC) sector is currently undergoing a digital transformation. The authors argued that this shift, driven by advanced technologies, is fundamentally reshaping the processes involved in designing, constructing, and operating built assets. These technologies encompass value-added monitoring of data from sensor networks, the implementation of robust semantic models for data management, and the integration of simulation and optimisation in engineering systems within the built environment. In the pursuit of decarbonising the built environment, Digital Twin offers a powerful approach, particularly in the operational stage. Through the utilisation of Digital Twin, comprehensive insights into the performance of buildings can be gathered, facilitating the identification of opportunities for substantial enhancements in both operation and comfort. For instance, the utilisation of sensors provides a mechanism to gather dynamic data, offering a wealth of information to make informed decisions. This information can be leveraged to capture, predict, simulate, and actuate, enabling a shift from reactive to proactive building management. This technological approach enables buildings to function with utmost efficiency from an environmental standpoint, thereby aligning with the objectives of decarbonising the built environment.

To enable the utilisation of Digital Twin, a series of processes related to data must be undertaken. These include the integration of diverse datasets originating from a variety of different sources, including sensors, controllers, internal databases, and external repositories. Furthermore, the development of homogeneous data models derived from raw data, along with a rigorous data cleaning process, is important to ensure the completeness and quality of the collected data. Collectively, these processes are referred to as data intertwining.

Various technologies are available for implementing data intertwining, each tailored to specific requirements. For example, Data Lakes serve as centralised repositories designed to integrate data from a diverse range of sources, and store data in its native format (i.e., structured or unstructured) [233]. In contrast, Data Warehouses, while also centralised, differ from Data Lakes in their approach to storing data, as they store data in a structured format [234, 235]. Additionally, Semantic data stores employ ontologies to establish relationships between distinct concepts within the stored data [236]. The selection of a particular technology for data intertwining hinges on a multitude of factors. These include the original format of the data, its quality and completeness, the degree of heterogeneity among various data sources, as well as the extent of required data pre-processing. For instance, in scenarios necessitating real-time predictive analytics, immediate access to processed data is imperative. Conversely, there exist instances where real-time access may not be as critical.

It is important to note that an exhaustive discussion of data intertwining technologies is beyond the scope of this research. Interested readers are encouraged to refer to the referenced works for a more comprehensive understating of these technologies and their applications.

2.5 Building Energy Performance

2.5.1 Overview

The building sector is a significant contributor to GHGs emissions, and roughly 40% of the EU's total energy consumption is attributed to buildings [237]. The International Energy Agency recognises energy efficiency as "the first fuel", in recognition of the multiple benefits of energy efficiency, including environmental and economic potentials [238]. This report also stated that the building sector has the highest unrealised energy efficiency potential, or more than 80%. While many governments have set policies to promote energy efficiency and reduce GHG emissions, building energy consumption is continuing to increase due to rising demand for building services, HVAC systems in particular, ensuring user comfort, along with the increase in time spent indoors [239]. In addition, there are other factors that influence energy performance, such as a building's design and characteristics, weather conditions, and occupant behaviour [240]. Despite many efforts to improve the energy performance, such as the European directive regarding the energy performance of buildings (EPBD) and the associated energy performance certifications (EPCs), buildings do not perform as anticipated. These instruments were established to improve building energy performance regulations among the EU member states and set binding goals that have to be translated into national regulations and energy policies [241].

Recent evidence has shown the limitations and shortcomings of energy certification schemes in many countries. For example, it was found that the input parameters (e.g., energy label, U-values) that are used in the Cantonal Energy Certificate for Buildings in Switzerland (CECB) are poor predictors of the actual energy use [242]. Moreover, a study conducted on the Irish EPCs concluded that reliance on the default values for thermal coefficients can lead to an overestimation of energy retrofit benefits [243]. Li et al. [244] conducted a comprehensive review of the EPC directives in EU member states and identified several issues in the current EPC, including the questionable reliability and credibility of EPC, the lack of representation of the entire building stock in EPC database, lack of performance monitoring , and input data quality. As a result, a discrepancy is found between the predicted energy performance as described by these certification schemes and the actual energy consumption, which is known as energy performance gap (EPG).

A growing body of literature has investigates the EPG and its underlying causes and solutions. EPG is defined as "the difference between expected energy consumption calculated by a building performance assessment and the actual consumption." [245]. Burman [246] identified three classifications of the performance gap as follows: 1) regulatory performance gap, which compares compliance modelling prediction (e.g., EPC modelling) with measured energy consumption; 2) static performance gap, which compares building energy simulation with actual operation conditions (i.e. performance modelling) with measured energy consumption; and 3) dynamic performance gap, which compares a calibrated performance model with measured energy consumption. Two classes of EPG have been suggested, namely perceived gap and actual gap [247]. The perceived gap occurs when the compliance modelling is conceived as the predicted performance and compared with the measured energy use, while the actual gap is a time-dependent gap that compares the measured energy use with the performance modelling predictions.

The underlying causes of EPG can be different for each building's life-cycle phase. A detailed examination of EPG in the pre-occupancy stage has identified the causes for the gap between building energy targets and the actual energy consumption during building commissioning phase [248], which include inadequate knowledge and collaboration between different stakeholders, as well as the lack of performance accountability during a building's use phase. In addition, [249] identified further causes, including design complexity and inaccuracy of design parameters, lack of post-testing and feedback, and lack of consideration regarding uncertainty. Building energy modelling uncertainty and inter-model variability that result from using different modelling tools have also been identified as contributing factors to the EPG [246]. In one of the most recent studies in this field, Cozza et al. [245] proposed two major categories for the causes of EPG, namely theoretical deviation causes and actual deviation causes. Under the first category, the authors identified several causes, such as the inaccuracy related to building modelling, occupant behaviour, and climate data. The second category includes causes related to sub-optimal building operation, malfunctioning equipment, measurement limitations, and execution of the work.

2.5.2 Energy Prediction Approaches

To reduce the negative environmental impacts of buildings during the operation phase, it is important to accurately measure the actual energy performance using a variety of techniques. Several studies have attempted to identify the different approaches of assessing building energy use (e.g., [240, 250, 251]). Seyedzadeh et al. [250] classified energy assessment techniques into four major classes: engineering method, simulation-based models, statistical methods, and ML-based modelling. Similarly, [240] suggested three main methods that are used for the assessment of energy use, namely: engineering method, statistical method, and ML method [240]. The engineering method is based on physics and typically conducted using simulation software, whereas the statistical and ML methods are data-driven approaches that seek to identify correlations between the outputs and input variables. Other researchers have classified these approaches into: numerical (i.e. simulation-based), analytical (i.e engineering method), and predictive approaches (ML models) [251]. Overall, these studies clearly agree on the main categories of building energy assessment approaches, despite their use of different terminologies.

A brief overview of each method follows. Engineering methods are based on using a building's physics and physical laws to calculate energy requirements at the building or system level. Solving the underlying equations requires extensive knowledge and understanding of a building's dynamics. Although developing these models requires significant effort, these models tend to be generalisable [251]. The simulation method is a computer-based model that can be developed using a variety of software tools (e.g., EnergyPlus, DOE-2, IES-VE, and TRNSYS). These tools are used for several energy-related applications, such as heating or cooling demands, lighting, and integrated energy system design. However, a comprehensive understanding and detailed information about the simulated system are required, including location and climate data, building operation schedules, construction components, and zones and surfaces. Although these simulation engines are effective and produce interpretable results, they tend to have low accuracy due to the lack of sufficient knowledge about the system's dynamics, especially during operation when occupant behaviours come into play [251]. In addition, developing these models requires considerable effort and is computationally expensive [240, 251]. In contrast, statistical and ML methods do not require a physical description of the building because the resulting models are purely data-driven. Essentially, these models utilise historical building data (e.g., energy consumption, weather data, indoor parameters, and operation schedules) to capture the underlying correlation between the desired outputs and the relevant variables [240, 250, 252]. This type of model is capable of predicting system behaviours under various conditions by training the models on a subset of the historical data. Once trained and calibrated, these models have the advantage of performing instantaneous prediction of building performance, which makes it desirable for real-time building control and optimisation [251, 252]. However, a key disadvantage of these models is the amount of historical data that are required to train the models, which also makes these approaches mostly applicable to existing buildings.

2.5.3 ML-based Energy Prediction and Optimisation

Several surveys of the literature have investigated the application of ML techniques for predicting the energy use of a building [250, 253–255]. ML-based solutions have been widely adopted due to the rapid development of ICT technologies (e.g., sensors, wireless transmission, and cloud computing), which facilitate capturing, storing, and processing of domain data [250], and due to their ability to solve complex and non-linear relationships that exist in multi-dimensional systems [254]. Mohandes et al. [253] found that ML models, ANNs in particular, have been applied to five main energy-related areas, namely: energy consumption, heating and cooling loads, indoor air temperature, HVAC systems, and heating and cooling of water systems.

Several studies have utilised a variety of ML algorithms—including ANN, support vector machines (SVMs), decision trees (DTs), and clustering—with diverse sets of input parameters in different energy-related applications. Petri et al. [256] utilised ANN and a multi-objective optimisation algorithm to optimise the set point for the inlet air temperature, thermal comfort, and energy consumption in a sports facility. Deb et al. [257] employed ANN to forecast diurnal cooling energy load for three institutional buildings and good prediction accuracy was achieved using energy consumption data for the previous five days. An ML-based solution has been proposed to predict heating energy demand for an institutional building using a variety of attributes, such as outdoor temperature, solar radiation, day type, occupancy profile, and characteristics of heating power level [255]. Li et al. [258] used SVM to predict the cooling energy load of HVAC systems in an office building in China. Their results show that SVM outperform ANN in terms of prediction accuracy. Although there is no systematic approach to select the most appropriate algorithm or input parameters, researchers often apply several algorithms and compare them based on prediction accuracy, computation time, number of required inputs and outputs, and the amount of data required to train the models.

While the applicability and significance of ML-based approaches for building energy prediction has been demonstrated, there is a relative lack of studies investigating the use of ML for the prediction and optimisation of energy in buildings with mixed-mode ventilation (i.e., a combination of natural ventilation and mechanical systems). Natural ventilation is an effective and sustainable building design option that has the environmental benefit of reducing energy consumption for indoor conditioning, while improving indoor environmental quality, including indoor air quality and thermal comfort [259, 260].

Several studies have shown a significant energy-saving potential as a result of using natural ventilation. For example, Tong et al. [261] estimated that up to 78% reduction of the cooling energy demand in office buildings can be achieved by harnessing natural ventilation strategies. Similarly, Barbadilla-Martín et al. [262] applied an adaptive control algorithm in mixed-mode office buildings equipped with HVAC systems and found potential energy savings in summer and winter seasons of 27% and 11.4%, respectively. Another study was conducted on mixed-mode school buildings in Spain to compare natural ventilation systems with mechanical systems in terms of energy savings, and the results showed that natural ventilation can reduce energy consumption by 18 to 33% [263]. Chen et al. [264] proposed an optimal control strategy using reinforcement learning to optimise HVAC and window operation. The reinforcement learning control reduced the HVAC energy use by 13% and 23% when compared to the heuristic control.

Most studies have been based on either simulation models, static data collected from the case study, or simulated datasets. Park and Park [265] argued that these approaches are insufficient to capture the dynamic aspects of natural ventilation systems and that the number of variables involved in modelling and evaluating the impact on energy consumption requires a more robust approach. These researchers experimented with a number of ML algorithms to predict the natural ventilation rate in an office space, including deep neural network (DNN), support vector regression (SVR), multivariate linear regression (MLR), and RF. The data acquired for these models are for indoor and outdoor variables, such as temperature, humidity, solar radiation, and pressure. The results show that DNN has the highest prediction accuracy, which can be applied to optimise windows control and operation. Another study developed a ML-based operation strategy, using the average daily outdoor temperature as a predictor, which predicted the best operating scheme of an engineered natural ventilation system [266]. Vrachimi et al. [267] proposed an ANN model to reduce the uncertainty of the airflow rate calculation by predicting local wind pressure coefficients for buildings of different shapes. The resulting coefficients significantly improved the simulated airflow rate when compared to an existing experimental database. Mousa el al. [268] developed a Classification and Regression Tree model (CART) to model the air change rate (ACH) in a naturally ventilated building. The authors obtained ACH values from a CFD simulation, while the input parameters were based on field measurements, including wind speed, direction, hour, and temperature. The latter input parameter was rendered non-predictive and removed from the input list. To avoid the high computational cost of conducting large scale CFD simulations (e.g. urban level), a study developed two data-driven models to predict a novel ventilation index, which links outdoor wind velocity to indoor airflow to assess the natural ventilation effectiveness of different design configurations [259]. The input data were based on CFD simulation and six design variables were employed in the prediction model (i.e., wind direction, relative sinuosity, building density, target building heights, height variation, and opening to wall ratio). Chen et al. [269] developed a prediction model for the thermal response of a room using the idea of pre-trained deep neural network (i.e., transfer learning) for the model's predictive control of HVAC and natural ventilation. The proposed model achieved high prediction accuracy for indoor temperature and relative humidity for several time intervals between 10 minutes and 2 hours. Gan et al. [270] proposed a deep learning model to predict natural ventilation potential (ACH in this study) by exploring the relationship between design features and indoor ventilation, as well as the outdoor airflow. The ACH prediction accuracy of the deep learning model was comparable to that of the CFD simulation.

Several indicators have been developed and introduced to measure the effectiveness of natural ventilation, including air change rate, airflow rate, and discharge coefficient [271–274]. The main difficulty with these indicators is that they require a thorough understanding of the physical model, which is mostly based on a simulation model and simulated data. Alternatively, several studies have used CO_2 concentration as a proxy to assess the adequacy of ventilation, and inform the operation and control of ventilation systems [275–278]. These studies have adopted a data-driven approach to predict the levels of CO_2 based on streaming data from sensors.

2.6 Discussion and Gap Identification

This review of the current LCA research landscape has identified several gaps and limitations, which will be described in this section.

2.6.1 Lack of Alignment with Domain Models and Manufacturing Systems

There is a growing interest in BIM-LCA integration. Nearly 20% of the analysed studies have used BIM, including using BIM as a source of data, a parametric model for energy consumption simulation, and as a simplified calculation tool for LCA by embedding environmental data into BIM objects. The most prevalent use of BIM in LCA applications is for geometric and material data acquisition. A similar finding was demonstrated by Potrč Obrecht et al. [143], who found that exchange of information is the most common link between BIM and LCA tools. However, this is currently limited by semantic incompleteness and interoperability issues between current software solutions. Soust-Verdaguer et al. [17] demonstrated that the BIM environment is missing a number of critical aspects that are important for environmental impact assessment, such as temporal processes, refurbishment and maintenance information, EoL treatment scenarios, and recycling data. Apart from the limitations in semantic information, automatic mapping of BIM data and LCA resources to facilitate the process of LCI building and resolve the interoperability issue is still lacking. In addition, a building's components and systems are produced through a manufacturing process. While the embodied carbon of the materials forming the final product is often fully considered, these manufacturing systems tend to not factor in the design configuration that is best conveyed via a BIM. In fact, automation of the building's production necessitates exploitation of information models (i.e., BIM) in each phase of the design, construction, and operation management life cycle. Current applications of BIM in projects mainly involve design information but often lack as-built and operation management information. Product manufacturers have recently engaged with BIM by making their manufactured products BIM compliant to enable designers to import virtual product specifications into their design environment, which provides a means to assess the environmental impact of their interventions.

2.6.2 Lack of Reasoning and Decision Support Capability

Buildings require a wide range of materials, products, and actors interacting in a dynamic and non-linear workflow as part of a complex ecosystem over the building's lifetime. Furthermore, there are many life-cycle variations that LCA tools and methods must take into consideration, including building usage, energy supply, changes in the energy mix, and occupancy patterns. Hence, minimising the environmental burden of buildings requires a comprehensive approach that factors in the complexity and the dynamic nature of building LCA. This requires the exploration of various scenarios to evaluate alternative design options, renovation strategies, and the generation of actionable improvement for building operations. In this context, the existing literature on decision support tools shows limitations in the proposed solutions, both in terms of scope and capabilities.

2.6.3 Limited Efforts to Scale Up LCA from Buildings to District Level

The literature related to the LCA of buildings at an aggregated level is scarce and has a multitude of heterogeneous methodological approaches [279]. There are two main approaches for building stock modelling, namely: a top-down approach that relies on some macroeconomic indicators, and a bottom-up approach that clusters buildings based on common characteristics [280]. In this review, it was found that the majority of studies focus on applying LCA on individual buildings or a group of buildings with complete background information about each building. Several studies have considered large-scale LCA applications for a variety of purposes, such as renovation of existing housing stocks [281], energy saving scenarios for EU-wide housing stock [282], and understanding the level of details required to conduct LCA at a large scale. The main challenge of scaling up LCA applications is that a trade-off must be made between the cost of collecting data and the reliability of LCA results. To deliver reliable and sound LCA results at district and city-wide level, it is essential to understand the level of detail required at the building level and the informative attributes at the district level. It is worth noting that a number of projects funded under the Horizon 2020 Smart Cities and Communities program are progressing the concept of Positive Energy Districts. These projects consider four dimensions in their district interventions, namely: energy efficiency, mobility, information and communication technologies, and citizen engagement. However, these projects fall short of embracing the LCA philosophy.

2.6.4 Lack of Support of Temporal Information

There is a need to factor in temporal information in the background and foreground LCI and LCIA methods to address maintenance, operation, deconstruction, and EoL treatment [5, 18, 19]. Construction processes involve longer time scales than are required in other industries [6, 17]. Therefore, considering the time dimensions in product system modelling is essential to understand the resulting pollutant emissions and resource consumption [5–7]. However, few studies have tested the use of dynamic data during construction and operation stages (11 studies only). Although electricity consumption is the main type of real-time data, some studies have used IoT devices for occupancy detection appliance use or have developed sensors to measure indoor temperature and relative humidity. Nonetheless, further research is required to determine the impact of accessing dynamic data and the frequency of data collection on assessment accuracy.

2.6.5 LCA: Directions for future research

The systematic literature review of LCA applied to buildings that was conducted in this chapter has highlighted key limitations and gaps of current LCA applications. There are three recurring themes in the gaps identified, namely semantics, temporality (i.e., dynamic data), and intelligence (to support decision making) as illustrated in Figure 2.8. These themes are applicable across the life cycle of a built asset, from concept design to EoL. The following subsections elaborate on each of these themes. LCA underpinned by semantics and informed by dynamic data can pave the way to a more accurate LCIA, while supporting decision-making, and the active control of buildings and districts. As such, there is a need to pave the way to a (near) real-time LCA capability that exploits a wide range of digital resources and which leverages intelligence (in the form of ML and optimisation algorithms) to assess the whole life cycle environmental impacts of built assets.

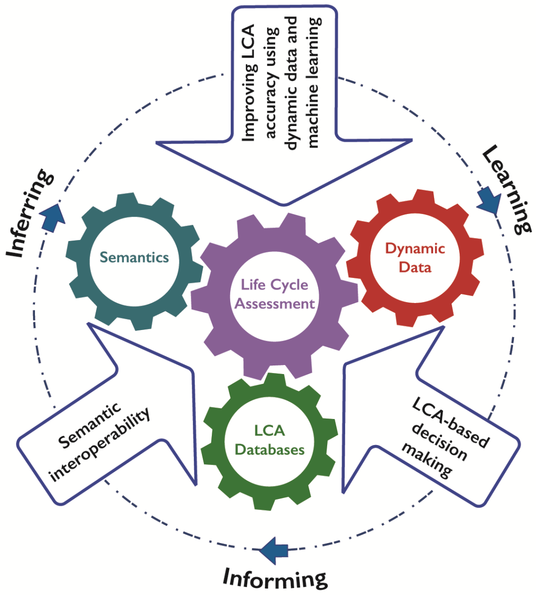


Figure 2.8: Key recurring themes in future LCA research applied to buildings

2.6.5.1 Semantic Interoperability of LCA

The concept of semantics refers to the reliance on computer-based models that provide a formal description of the context that underpins the domain under investigation [283]. In practice, the domain conceptualisations that are held by stakeholders and software across disciplines tend to be incompatible and necessitate ad hoc solutions [284]. Furthermore, the use of semantics, including BIM and GIS, provides a means to integrate and contextualise existing inventory databases, and provides a sound basis to streamline the LCA process of buildings and districts. However, this will require an inventory of existing LCA databases, methods, standards, and tools to be established. In addition, their underpinning semantics should be elicited. Furthermore, the existing relevant semantic models, such as BIM and GIS, and current LCA databases should be expanded to address the completeness requirements that are necessary to provide holistic accounts of the environmental impacts of buildings and wider districts. The key methodological challenges in delivering semantic LCA require a comprehensive (life cycle and supply chain) understanding of the semantic resources that are required to deliver life-cycle assessment at building and district level. A reference architecture for semantic LCA that factors in existing databases, models, methods, and tools is also required. Finally, a consensus and requirements of semantic and dynamic LCA should be developed.

2.6.5.2 LCA Based on Dynamic Data

Research is needed to assess the impact of utilising dynamic data on the accuracy of LCA results throughout different project stages, such as construction and operation. Delivering real-time accounts of the life-cycle performance of buildings and districts requires multi-aspect sensory data, including: (a) indoor and outdoor environmental data, and (b) building and district performance data (e.g., energy consumption, pollution, and carbon emissions). The collection of dynamic data will require the identification of necessary instrumentation and data capture technologies, while leveraging existing building management systems and Information and Communication Technology (ICT) infrastructure. This requires a context to the sensed data to be provided via semantics. In addition, a systems approach should be adopted, whereby the performance and environmental impact of a physical artefact (e.g., a built asset) involves the assessment of each constituent subsystem.

2.6.5.3 Machine-Learning-based Decision Making

Research is required to evaluate the impact of semantic and dynamic LCA in the decision-making process by non-experts, which should explore a wide range of options and scenarios with the least environmental impacts, while also advising on corrective measures through actionable ML. In addition, ML techniques, including model predictive control and optimisation algorithms, can be used to deliver actionable knowledge to inform various control strategies and corrective actions with a view to reducing the gap between the predicted and actual environmental impacts. They may also be used to overcome data gaps for ML. In addition, ML technologies may be used in real-time applications to monitor and control the systems in a way that reduces negative environmental impacts. ML models may be more easily integrated than other black box methods because they are more easily interpreted by the users. However, the monetary and time costs of establishing ML models should be considered for real-time use.

2.7 Conclusion

The primary focus of this chapter was to address research question 1, which asks:

What are the key limitations of current LCA methods that affect the accuracy and widespread adoption of LCA in the building domain?

This chapter has mainly presented a review of the research progress in the field of building LCA, focusing on the current applications of LCA in buildings. In addition, this chapter has highlighted gaps and limitations of the current LCA applications in the building domain.

There is an increasing adoption of building LCA across the life-cycle stages of a building, including manufacturing the building materials, design, construction, use phase, and EoL. However, successful LCA implementation must factor in the dynamic nature of buildings, variable operational and environmental conditions, the long time scale of buildings, and the specific challenges that are associated with each life-cycle stage. In particular, the challenges associated with LCA in the operational stage stem from several factors, including a variation in operational energy demand, energy system evolution, building use and occupancy patterns, and building and environmental regulations.

This review of LCA applied to buildings has revealed several research gaps and limitations, including the lack of alignment with domain models and manufacturing systems, lack of reasoning and decision support capability, limited efforts to scale up LCA from buildings to district level, and lack of support of temporal information.

While previous efforts have led to incremental progress, this study will explore the concept of semantics to integrate and contextualise the existing domain models (e.g., BIM), LCA tools, and inventory databases to streamline the LCA process and provide holistic accounts of the environmental impacts of buildings. This thesis also intends to develop a decision-support system that leverages dynamic data, ML, and optimisation methods for real-time assessment, monitoring, optimisation, and control of buildings during the operation stage. These topics will be examined in greater depth in the following chapters of this thesis.

Chapter 3

Research Design and Methodology

3.1 Research Methodology

This chapter explains and justifies the research methodology that was adopted to deliver this research. It starts by discussing the philosophical stances of scientific research and aligning the current study with the relevant school of thought. Following this, the research approach is presented, which will be used to elaborate on the research questions that were posed in Chapter 1, and the approaches undertaken to address them will be described. The main goal of this chapter is to provide a holistic view of this thesis by linking the various chapters, research questions, and employed techniques.

3.1.1 Theoretical Background

Research methodology can be broken down into two components: the first is research, which is defined as "a quest for knowledge through diligent search or investigation or experimentation aimed at the discovery and interpretation of new knowledge" [286], and the second is defined as "a systematic body of procedures and techniques applied in carrying out investigations or experimentation targeted at obtaining new knowledge" [286]. Research philosophy serves as the foundation for research by defining the nature of reality (i.e., ontology), sources of knowledge (i.e., epistemology), and the role of beliefs and values (i.e., axiology) [287].

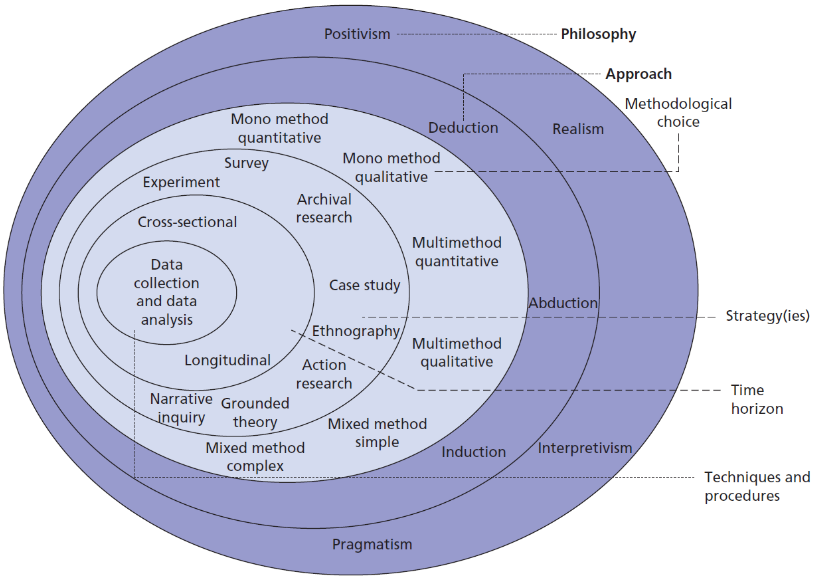


Figure 3.1: The research 'onion' [285]

Guban and Lincoln argue that *"Questions of method are secondary to questions of paradigm"* [288], and emphasise the importance of research paradigms in the discovery and creation of knowledge. Although many categorisations and classifications of existing research philosophies (e.g., Guba and Lincoln [288], Saunders et al. [285], Ritchie et al. [289]) are found in the literature, there is no consensus on the implications of each philosophical stance on the selection of available research methods [290]. Furthermore, as noted in [290], scholars of research philosophies have used contradicting terminologies, even when referring to the same concept; for instance, using the term *'approach'* instead of *'method'* to describe a specific methodological choice (as presented in Figure 3.1).

The main objective of this chapter is to elaborate on the research methodology that was developed during the course of the current research rather than debating the scholarly work related to research philosophy. Therefore, this section will only outline the basic tenets of research methods because a comprehensive review is beyond the scope of this study. Therefore, the model and the terminologies proposed by Saunders et al. [285] will be followed for the sake of consistency. The following sections start by discussing the outermost layer, the research philosophy (see Figure 3.1), which will be followed by a description of the overall research stages and the associated strategies and processes.

According to Saunders et al. [285], the choice of research philosophy (i.e., philosophical stance) has a profound effect on all of the subsequent steps throughout the research development, including the underpinning research strategy, methods, and decisions related to data and analysis techniques. There are four distinct philosophical stances, namely positivism, interpretivism, realism, and pragmatism [285]. Positivism is the philosophical stance of the natural scientist, which involves data collection of an observable phenomenon to understand causality and provide generalisations in an objective manner [285], such that the researcher remains objective and independent of the data and that the research findings are inferred empirically. Hence, this stance often employs quantitative methods.

Realism is similar to positivism in the sense that both follow a scientific approach during data collection and interpretation of the data [285]. However, realism holds that reality exists independently of human perceptions. Two schools of thought can be found within this philosophical position: direct realism, which argues that humans experience the real world through sensation; and critical realism, which suggests that interpretable sensations may not be the real objects of the world [285].

Unlike positivism, interpretivism asserts that insightful concepts are lost when the world's complexity is limited to law-like generalisations similar to those found in physics [285]. In this philosophical stance, reality is seen as a social construct that is interpreted by social actors or individuals. Hence, interpretivism is subjective in nature, in which the researcher's perspective is an integral part of the qualitative research methods that are used with this paradigm.

While these philosophical positions have clear views of the reality and specific methods to answer research questions, pragmatism uses multiple reality and mixed research methods in the development of knowledge. Pragmatism is particularly appropriate for this current research for several reasons. First, Saunders et al. [285] assert that the research question is a crucial factor in choosing the philosophical stance. In this regard, answering the research questions posed in Chapter 1 requires subjective and objective methods. Therefore, with this paradigm, it is possible to apply multiple philosophical stances in the same research. Moreover, an important feature of pragmatism is that it is *'a social model of knowledge'*, which recognises the fact that knowledge is a social and not an individual achievement because the idea is only valid when people in a specific domain share the same perception [291]. This last point is particularly relevant to the current research given the involvement in a research project, during which explanation, interaction, and dialogue with experts have a significant role in the selection of research methods and analysis techniques. In addition, the current study was carried out through multiple stages, where each stage requires different research methods, be it qualitative or quantitative. Action research is associated with the qualitative aspect of this research, which informed the selection of the research methods throughout the research development. The quantitative aspect of this research was carried out during the development of the data-driven use case and the validation of the proposed solution on a real case study.

3.2 Research Design

This section will describe the research design, which was broken down into three stages (as illustrated in Figure 3.2). The first stage was an exploratory exercise by way of a literature review to identify the existing applications, and identify limitations and gaps in the existing body of research to inform and refine the research questions. In the second stage, a participatory action research strategy was carried out through engagement with stakeholders and industry experts, and by contributing to a research project. In the final stage, the lessons learned and the insights from the previous stages informed the development of the analysis and validated the main research objectives.

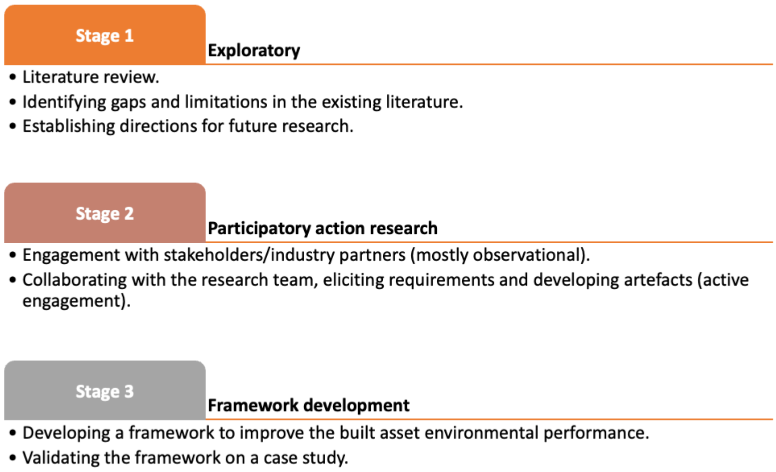


Figure 3.2: Summary of the research design

3.2.1 Stage 1: Exploratory

The primary activity of Stage 1 is to carry out a literature review to theoretically analyse the existing body of knowledge pertaining to the environmental performance of buildings. This exploratory step serves many purposes, including recognising the state-of-the-art research of LCA applied to buildings, identifying the shortcomings of proposed solutions, identifying the requirements and functionality needed to streamline LCA throughout the various life cycles of the asset, and most importantly support the refinement of the research questions. The literature review also formed the foundation for the identification of LCA use cases across the physical, temporal, and technological dimensions. In particular, this stage highlighted the areas that can considerably improve LCA practice and reliability in the construction industry by adopting the concept of semantic interoperability to integrate various artefacts and utilising dynamic data and decision support systems that use ML and optimisation algorithms to improve the environmental performance of built assets.

3.2.2 Stage 2: Participatory Action Research in the LCA of Buildings

Stage 2 uses a participatory research approach via engagement in the research project, namely **SemanticLCA**. Essentially, participating in a research project can be considered as an experiential learning experience, as Kolb defines it *"the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience"* [292]. Initially, the involvement in the research project was observational to understand more about the domain and the underpinning concepts. The project then gradually contributed to the deliverables and engaged with industry experts and facility managers. Presenting conceptual models and solutions to the project's participants provides important feedback to improve the initial solutions and identify the challenges during the implementation and testing. Several workshops took place with domain experts from the UK and across Europe to discuss the current LCA practice, and to identify use cases to be developed and investigated.

3.2.2.1 SemanticLCA Research Project

SemanticLCA is a collaborative research project between Cardiff University and the Luxembourg Institute of Science and Technology (LIST), which is funded by the Engineering and Physical Sciences Research Council (UK) and Fonds National de la Recherche (Luxembourg). This research project aims to promote scalable, cradle-to-grave environmental sustainability capabilities by leveraging building and district semantics to streamline LCA and to devise corrective actions. Several objectives were established to achieve this overarching goal, including: i) establish an inventory of the existing life-cycle assessment databases, methods, standards, and tools and then elicit their underpinning semantics (i.e., data structures); ii) develop a reference architecture for Semantic LCA, which will leverage existing information sources, including BIM (IFC), GIS (CityGML), and current LCA databases to address completeness requirements necessary to provide holistic accounts of environmental impacts of buildings and wider districts; iii) expand or align existing relevant semantic models, including BIM and GIS, to factor in dynamic data and deliver a dynamic life-cycle assessment capability of buildings and districts; and iv) develop ML techniques, including model predictive control and optimisation algorithms, to deliver actionable knowledge, and to inform various control strategies and corrective measures with a view to reducing the gap between predicted and actual environmental impact. Engagement in this project has been a valuable source of experience at the individual level and has guided the formulation of use cases that will be used to validate the proposed methodology.

During the course of the research project, different engagement modalities were employed, including workshops and periodic meetings. Two workshops were held, bringing together academics, industry partners, and the research team. These sessions started with an introduction to the research project, followed by the presentation of preliminary research findings. Then, a series of use cases were deliberated upon, inviting discussions regarding their validity, and practical implications. These workshops provided invaluable high-level feedback and recommendations for refinement. In addition, regular periodic meetings were conducted on a weekly and monthly basis. These meeting served as platforms for in-depth discussions on research progress, assessment of existing LCA tools and frameworks, and evaluation of LCA use cases development and proposed solutions.

3.2.2.2 Requirements Elicitation

The procedure that is adopted in this stage of scoping and developing the SemanticLCA system aims to formalise LCA use cases identification, specify the requirements for the identified use cases, and design and develop an overall semantic platform. The procedural approach undertaken in this stage of scoping and formalising LCA use cases was not predetermined, but rather evolved organically in response to the iterative nature of the participatory action research. The steps involved can be viewed as a post-rationalisation of the procedure pursued.

Use Cases Identification

The research initially aimed to explore the landscape of LCA applications within the building domain. From the literature review, it became evident that there is a wide range of applications for LCA throughout the different stages of a building's life cycle. In this regard, developing a taxonomy can be a highly effective approach for understanding a complex domain like LCA in buildings. Essentially, a taxonomy allows for a systematic and comprehensive understanding of the various facets and applications within the domain. It provides a structured framework for categorising and organising key concepts and use cases, which is particularly valuable in a multifaceted field such as LCA. While it is true that other approaches such as surveys, expert interviews, and data mining could have been considered to achieve a more comprehensive understanding of the domain, it is crucial to acknowledge that the selection of approach was constrained by the specific context and available resources within the research project.

Semantisation of Use Cases

The second step of the elicitation process identified the key concepts that are required to implement the use cases. To achieve this, a semantisation of use cases technique was developed through weekly engagement with the research team, by which a use case can be analysed from different perspectives. The development of this technique did not follow a specific methodological approach; rather, it emerged from a synthesis of collective knowledge, a literature review, and a process of brainstorming. These combined efforts facilitated the identification of pertinent concepts and dimensions essential for the semantisation of the identified LCA use cases. The developed technique serves two main purposes: first, identifying the characteristics, objectives, data sources, and modelling technique for each use case; and second, establishing a generic architecture, including its components and processes, to realise the identified use cases. This technique will be utilised in Chapter 6.

Analysis of the Use Case's Processes and Mechanisms

The research project involves the development of a software process, wherein diverse software components are integrated. These include BIM, energy simulation software, and Brightway2 for LCA modeling. Additionally, the devised solution leverages data sourced from sensors, machine learning and genetic algorithms for prediction and optimization. In the realm of software development, a critical aspect entails the consideration of integration and the flow of information between these entities (i.e., objects). In this context, Unified Modeling Language (UML) stands out as the widely acknowledged standard in software development [293]. UML is a standard language that was started in 1994 and is used to document, visualise, and specify the development of a software-intensive environment [294], such as conducting LCA of buildings.

UML has a plethora of modelling concepts and diagrams that are used throughout the life cycle of a software system. The 'Sequence diagram' is relevant to the current research approach because it has been used in semantic modelling [295]. Sequence diagrams are a subcategory of interaction diagrams that are concerned with the interaction and the time ordering of communications between a set of objects or artefacts. In the context of LCA, the objects are (for example) a BIM model, LCA software, and sensors (Figure 3.3). The interactions between these objects, or messages that an object sends or receives represent the exchange of information between the objects, which can identify the relationship between two objects. These messages between objects are represented by horizontal solid or dashed lines. Another important feature of the sequence diagram is the vertical line that is associated with each object. This line indicates the time horizon of the object interactions.

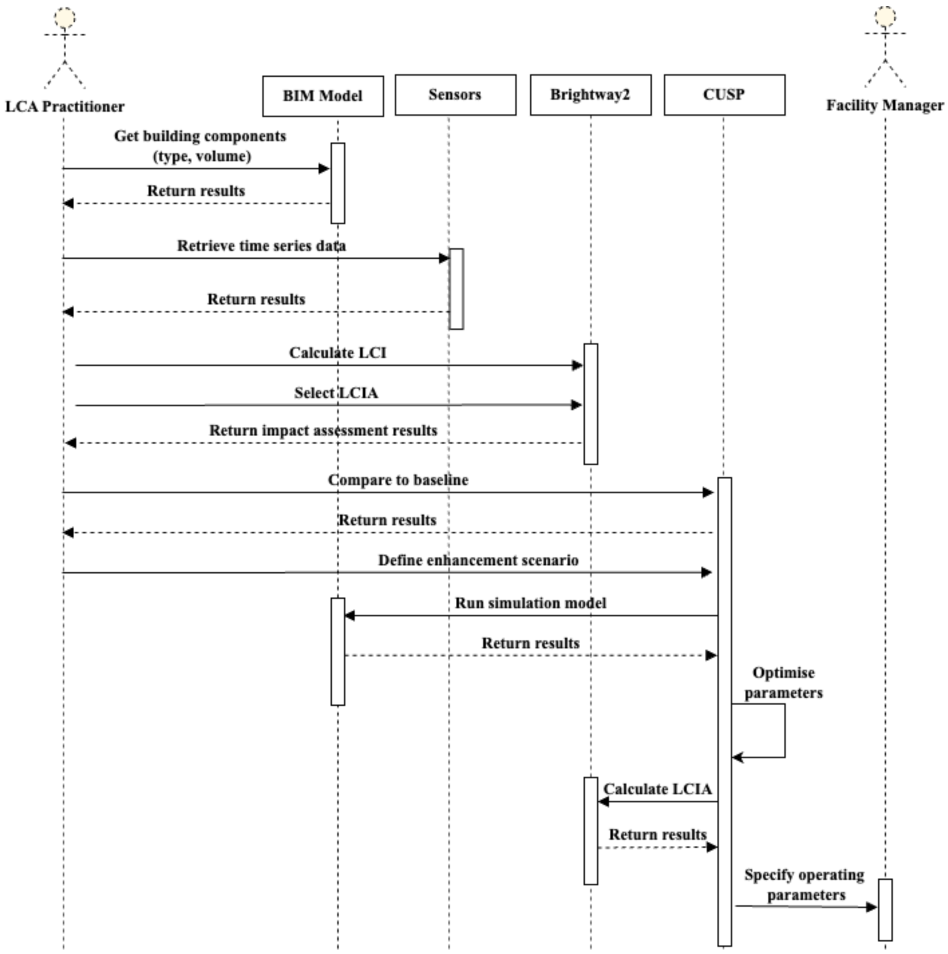


Figure 3.3: Simplified example of a sequence diagram for an LCA use case

3.2.3 Stage 3: Framework Development

Stage 3 represents the core contribution of this study, which will focus on a specific use case based on a framework developed during this stage and informed by the literature review. As presented in Chapter 2, the energy performance gap that is found between the predicted energy consumption in the design stage of buildings and the actual energy use during operation underestimates the environmental impacts of buildings during the use phase. In addition, the majority of LCA studies have considered static assumptions. In other words, static LCA does not consistently factor in the temporal variations of a building, such as building usage and indoor conditions. The other problems pertaining to the last point are as follows. First, building operation must be considered as a multi-faceted problem, in which facility managers attempt to maximise indoor comfort and indoor environmental conditions, and simultaneously minimise energy consumption. Second, establishing the life-cycle inventory in a dynamic manner can be challenging from a technical point of view because the data sources that need to be consolidated come in different data structures and with different levels of granularity. Therefore, aggregating these data is time consuming and computationally expensive. While previous efforts have led to incremental progress on multiple fronts, they lack a holistic approach that takes into account the integration of different domain models and data sources, the consideration of multiple objectives (often conflicting) during the operation of a building, and the development of a decision support system to help contextualise and translate information into actionable measures. The outcomes of this stage will be the subject of Chapters 4, 5, and 6.

3.3 Overall Methodology

This section will provide an overview of the methodology used to achieve the thesis objectives. The content will include a general description of the overall approach for addressing the hypothesis and research questions. The methodology encompasses three interrelated aspects: a framework, a use case, and a case study (Figure 3.4). Additionally, the justifications for the decisions and choices made will be discussed.

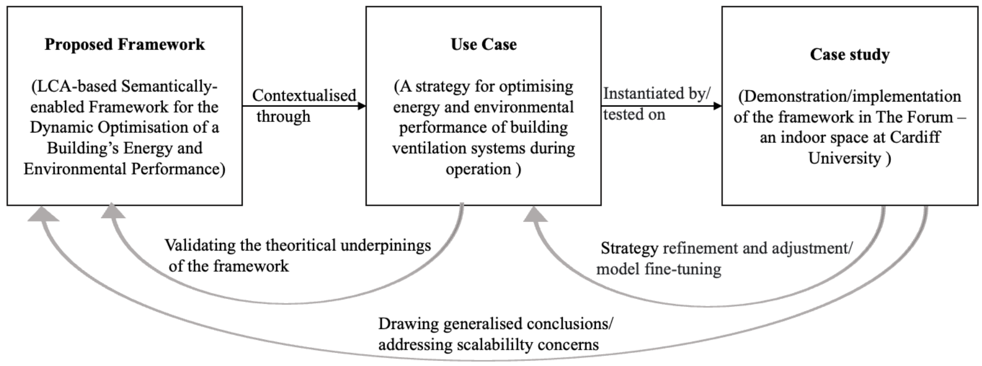


Figure 3.4: Simplified representation of the overall methodology

3.3.1 The proposed Framework

In the pursuit of advancing sustainable practices in the building domain, this study employs a semantically-enabled framework integrating LCA, dynamic data, ML and optimisation techniques, and digital resources (e.g., BIM). This framework is in direct alignment with the research questions and the central hypothesis that posits "*a semantic-based approach can facilitate the process of LCA and improve the accuracy of the LCA results by leveraging the value of dynamic data, learning systems, and digital built-environment resources*".

The incorporation of ML in the framework is substantiated by the growing tendency to utilise ML for various LCA applications [196]. A recent review indicates the potential of ML as a valuable tool in optimising and streamlining the process LCA scenarios [195]. Furthermore, the integration of ML with traditional optimisation methods presents a promising avenue for expediting the exploration and evaluation of different scenarios [198]. By integrating both ML and Optimisation techniques within the framework, this study leverages the strengths of each component to address the complexities and challenges inherent in optimising building energy and environmental performance during the operational phase.

The inclusion of dynamic data is imperative for achieving accurate and reliable LCA results, particularly due to the inherently dynamic nature of buildings. Real-time data, sourced from various sources such as smart utility meters, Internet of Things (IoT), as well as sensors for monitoring indoor conditions, plays a pivotal role in improving LCA accuracy [188–190].

The use of semantics in the framework is crucial due to the interdisciplinary nature of LCA, which requires integration across diverse fields and data sources. Semantics, facilitated through ontology and semantic modelling, addresses challenges in data interoperability, and information sharing [206]. While prior studies demonstrate the potential of ontology-driven approaches to overcome limitations in existing data format [221], LCI modelling [224], and address the impact of spatiotemporal scopes on LCA results [223], further research is warranted, particularly in areas like the building domain, where ontology-based approaches have yet to be extensively explored.

3.3.2 Identified use case

Use cases provide a context within which the theoretical framework is applied, address specific challenges or problems, and serve as a practical demonstration of how the framework is deployed in a real-world scenario. Insights gained from the practical application of the use case serve to validate the theoretical underpinnings of the framework, ensuring its applicability in real-world scenarios. This can potentially lead to an extension of the framework beyond the specifics of the use case.

In this study, the identified use case seeks to reduce the environmental impacts of a building's energy consumption during the operation phase, with a focus on the mechanical ventilation system. This is achieved by comparing two scenarios: a baseline scenario that represents a schedule-based, static operation strategy for the mechanical ventilation system; and an optimised scenario, which is developed using ML and optimisation techniques that take the dynamic indoor conditions captured by indoor sensors into account, such as CO_2. This use case was selected after thorough discussions with the project team in Stage 2. It emerged from the application of the developed LCA use cases taxonomy (Appendix A). Moreover, this choice directly addresses a significant challenge in the building domain highlighted in Chapter 2 — the energy performance gap [245, 248]. This gap underscores the underestimation of a building's environmental impacts during its operational phase.

3.3.3 Research Case Study

The case study strategy evaluates a research problem or phenomenon under certain circumstances or real-life scenarios [285]. Myers and Avison [296] state that in a case study, there is no experimental control of contextual variables, and the examination of a phenomenon is carried out in a natural setting. Hence, this would lead to a deep understanding of the research problem and may potentially increase the validity of the results [285]. Furthermore, the insights derived from observations and empirical findings in the case study play an important role in refining the use case strategies, selection of parameters, and fine-tuning of applied models. This iterative process ensures the use case can effectively achieve its objectives. Also, the feedback from the case study enables the extraction of generalised conclusions that are instrumental in addressing scalability concerns, particularly in the application of the framework across larger-scale environments such as an entire building or multiple buildings.

The case study was based at the Queen's Buildings, which is a complex of connected buildings at Cardiff University (UK). Figure 3.5 shows a 3D model of the buildings that was developed using Autodesk Revit. In general, the buildings are naturally ventilated except for a few spaces where mechanical ventilation is installed to supply fresh air (the buildings have no cooling functionally). The current study is based on one of those spaces that has a mixed mode ventilation (i.e., natural and mechanical ventilation), which is referred to as 'The Forum' and is located on the first floor of one of the buildings (the west building). The Forum has a total area of 323 m^2 with a design capacity of 200 people. The Forum is an informal meeting area, and is typically occupied by students and staff of the School of Engineering. The mechanical ventilation system has several supply diffusers and extraction grills distributed throughout the space. The natural ventilation has 13 east-facing windows and two south-facing windows. This space was selected based on the following criteria:

- The Forum has mixed-mode ventilation, which allows for a comparison and optimisation of multiple scenarios.
- Due to the functional characteristics of the space, it is anticipated that considerable degrees of variability in indoor conditions and noticeable fluctuations in the occupancy profile will be observed.
- The researchers have unrestricted access to the space to maintain the sensing infrastructure or to carry out any necessary repairs.

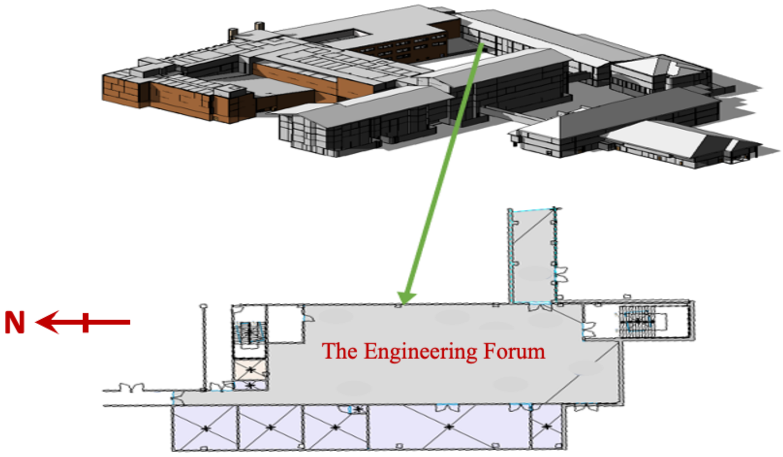


Figure 3.5: BIM model of the demonstration building and the floor plan of studied area.

3.4 Environmental Monitoring System

This section describes the data collection infrastructure, which uses an IoT-based indoor environment monitoring system and an on-site weather station to collect outdoor environmental conditions. The indoor monitoring infrastructure is a customised low-power IoT system that has three main hardware components, namely an end device (i.e., remote unit), a gateway, and a server. Data are communicated between the network components via alow power wide area networking (LoRaWAN) protocol (Figure 3.6). The remote unit have several integrated sensors. Each sensor collects data for a certain indoor parameter. The sensed data are transmitted to a gateway within the same vicinity. The gateway then forwards the data to a network server that processes and stores the data for later analysis.

A description of each sensing device or equipment acquired in this research follows. These sensors can be divided into three categories: indoor environment sensors, window sensors, and weather variables sensors. Table 3.1 also provides a brief summary of all of the sensors regarding their category, model, and the associated

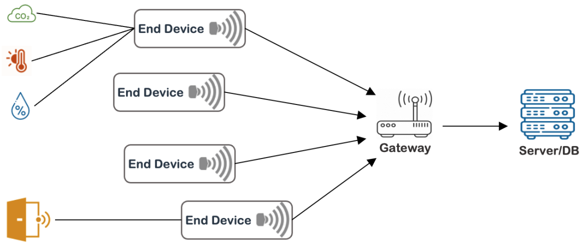


Figure 3.6: Schematic diagram of the indoor environment monitoring system parameter.

Table 3.1: Summary of the technical specifications of the employed sensing devices.

Variable(s)	Sensor	Measurement range	Accuracy
CO_2	Sensirion (SCD41)	400 to 5000 ppm	$\pm$ 40 ppm
Temperature	Bosch (BME680)	$-$ 40 to 85 $°C$	$\pm$ 0.5 $°C$
Humidity	Bosch (BME680)	0 to 100 %r.H	$\pm$3 %r.H
Pressure	Bosch (BME680)	300 to 1100 hPa	$\pm$ 0.6 hPa
Particulate matter	Plantower (PMS5003)	0 500 $\mu g/m^3$	$\pm$ 10 $\mu g/m^3$
Window status	Dragino (LSS02)	0/1	-
Weather data	Davis (Vantage Pro2)	Follow the link [1]	-

Indoor environment sensor: The current research used a combination of different sensors to measure several indoor parameters, including temperature, humidity, CO_2 level, and particulate matter. These sensors were integrated

[1]https://www.davisinstruments.com/pages/vantage-pro2

in a single box, which was referred to as the 'end device', as shown in Figure 3.6.

In this study, a single remote unit was used and placed at a height of 2.5 meters above the floor in the centre of the Forum. This height was selected to mitigate the risk of tampering, given that the space is used by engineering staff and students. While the remote unit position deviates from some established standards and guidelines, such as ASHRAE Standards 62.1 [297], which recommends placing CO_2 sensors in the breathing zone at a height of up to 1.8 meters above the floor, the decision regarding the position was primarily made to prevent any damage to the remote unit.

- **CO_2 sensor (Sensirion SCD41):** This sensor measures the CO_2 concentration in the air using photoacoustic non-dispersive infra-red technology. Essentially, CO_2 molecules absorb the energy emitted by the infrared, which causes the molecules to vibrate inside the measurement chamber. This vibration creates acoustic waves, which are captured by a microphone, by which the CO_2 level can be calculated.
- **Temperature sensor (Bosch BME680):** The indoor temperature is measured based on the voltage change of a silica diode-based temperature measurement. The working principle of this type of sensing device is that the voltage across a diode changes in response to the increase or decrease in the temperature of the surrounding air.
- **Humidity sensor (Bosch BME680):** This sensor uses a capacitive humidity sensor to measure the presence of water vapour in the air. The humidity level is measured by the relative electrical capacity changes of a polymer-based capacitor.
- **Pressure sensor (Bosch BME680):** The indoor pressure is measured based on the deformation of a highly sensitive thin membrane in response to changes in atmospheric pressure.

- **Particulate matter sensor (Plantower PMS5003):** This sensor measures the level of two types of particulate matter in the air (i.e., pollutants), namely PM10 and PM2.5. The presence of a certain particle in the air is detected by injecting a light source, typically a laser beam, through the air sample to measure the scattering of the light, which can then be translated into a mass concentration of particulate matter (μg/m^3).

Window sensor (Dragino LSS02): This sensor detects window open/close status using a magnet and a reed switch, whereby the electric current in the sensor is disrupted when the magnet is far apart from the sensor, indicating an open status, and vice versa. Data are transmitted to the gateway for each open/close event. The duration of each status is calculated by taking the difference between the timestamps of two distinct actions.

Weather data (Davis Vantage Pro2): An industrial-grade weather station was installed on the roof of the west building. The weather station integrates a plethora of sensors to collect real-time data for more than 20 variables, including temperature, humidity, pressure, precipitation, wind direction and speed, solar radiation, and so on.

3.5 Summary

This chapter has presented the research design and methodology of the current study. This chapter began by describing the philosophical aspects of the research paradigm adopted for this research, which was followed by a high-level description of a multi-stage research design. These stages include an exploratory stage that was based on the literature review to identify gaps and limitations in the existing body of knowledge; a participatory research stage, in which engagement and collaboration with other researchers and industry experts identified the shortcomings of current practices and provided feedback on a candidate framework; and the final stage applied the learning outcomes from the previous stages. The overall methodology was then discussed, including a description of the framework components, identified use case, demonstration site, and data collection methods. The third stage will be the subject of the remaining chapters of this thesis.

Chapter 4

LCA-based Dynamic Environmental Performance Framework for Buildings

As described in the previous chapters, the present study aims to develop a semantic-based framework that enables the delivery of near real-time environmental footprint assessments to support effective operation and management strategies for the built environment. The proposed methodology incorporates dynamic data, learning systems, and other digital resources to facilitate the application of LCA. To validate this approach, a use case was derived with the aim of optimising the environmental performance of buildings during the operation stage.

In this study, the notion of "dynamic" can be seen from three distinct perspectives. First, the employment of dynamic data, particularly sensor-generated data and energy consumption data, allows for a comprehensive understanding of a building's operational needs and its inherent dynamics. The dynamic data will be contextualised and subsequently employed to dynamically enhance operation strategies. Second, the proposed framework enables a dynamic and demand-controlled operation strategy. This departure from conventional fixed-schedule strategies is pivotal. The framework provides the capability to promptly respond and adapt to changing environmental conditions. This not only enhances energy efficiency but also contributes to a more sustainable and environmentally conscious operational model. Third, fundamentally, the distinctive feature of the proposed framework, as compared to traditional LCA (i.e., static LCA), lies in its consideration of dynamic variations. Unlike traditional LCA which relies on generic data derived from typical building usage and operation, the proposed framework accounts for these dynamic variations. This results in a more accurate and granular assessment of a building's environmental impact. The framework enables continuous calculation of environmental footprint using learning systems grounded in factual data. These perspectives collectively push the proposed framework towards a more advanced and dynamic approach to optimising building energy and environmental performance.

The findings are presented in three interlinked chapters, which provide a comprehensive overview of the methodology and its implementation. Chapter 4 outlines the conceptual framework that was developed to deliver a near-real-time environmental assessment. This includes the use of dynamic building data, such as energy consumption and indoor conditions data, as well as a number of modelling techniques, such as simulation, prediction, and optimisation. These topics are discussed in detail in this chapter. Chapter 5 presents the output and results of implementing the framework, which includes an LCA-based assessment of the environmental impact of different operation scenarios. Finally, Chapter 6 will discuss the semantic modelling that is used for the proposed framework. The deferral of semantic modelling to Chapter 6 is grounded on the fact that a thorough understanding of the framework is a prerequisite for informed semantic modelling. Before embarking on the semantic modelling of the framework, a comprehensive understating of the framework, its information exchange dynamics, and the interplay of its components is essential.

4.1 Framework Components

The proposed framework for delivering the use case has several interconnected components, as presented in Figure 4.1. The main components of the framework include data sources, building energy simulation, an ML-based prediction model, an optimisation model, and an LCA model. The data sources include sensor data, weather data from the on-site weather station, and the BIM model. The building's energy simulation will be used to generate data for the studied scenarios. The simulation data, along with data from other sources, will feed into the ML-based prediction model, which will have several outputs, including energy consumption and CO_2 concentrations. The outputs of the ML model will be used as inputs to the optimisation model, which aims to minimise the energy consumption of the building system. Finally, an LCA model will be used to compare the environmental performance of the developed scenarios.

The following sections aim to provide a technical background for the employed components of the framework, which are critical for the delivery of the use case. First, a detailed description of the ML-based prediction model will be given, which utilises the collected data to predict various outputs, such as energy consumption and CO_2 concentrations. Subsequently, the optimisation algorithm that is used in the framework will be introduced, which aims to minimise energy consumption while maintaining a satisfactory indoor environment. This will be followed by an overview of the energy simulation, which generates data for the scenarios that will be studied. Following this, the LCA method, which is used to evaluate the environmental performance of the developed scenarios, will be discussed. Finally, an explanation of the nature of the data collected for the study will be presented, including the considered parameters, data preparation, and preprocessing. It is worth noting that the exact configuration of each model to generate the required outputs for the use case will be discussed in the subsequent chapter (Chapter 5). These sections solely intend to provide a technical understanding of the utilised models.

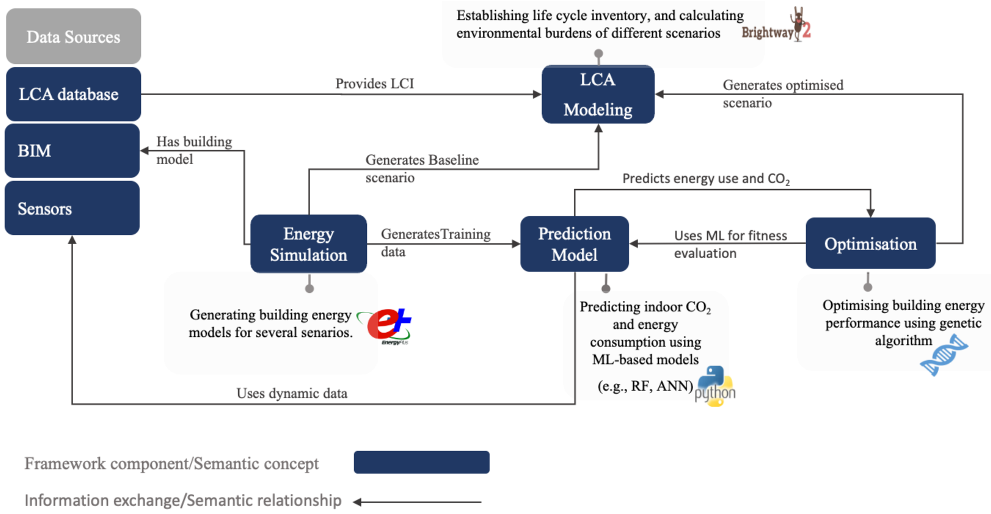


Figure 4.1: Visual representation of the framework used to deliver the use case showing key components and workflow.

4.2 Machine Learning

While the use of ML extends across a broad range of applications, it's imperative to note that in the context of this thesis, ML techniques are being employed for a specific purpose. As mentioned in the previous chapter, the framework's focus lies in its application to a specific use case. Within this use case, the primary aim is to predict two distinct parameters: the concentration of CO_2 in indoor environments and the energy consumption of mechanical ventilation systems. Predicting CO_2 levels serves two purposes: first, to ensure that the space maintains recommended CO_2 levels, which is essential for providing a healthy indoor environment; and second, it enables the determination of the ventilation requirements of the indoor space, which facilitates proactive control of the air supply (either through mechanical or natural ventilation) to increase energy efficiency and reduce the environmental impact of building operations. In addition, the prediction of CO_2

and energy consumption will be used to develop the optimal operation strategy for the ventilation system.

4.2.1 Feature Selection

The primary purpose of feature selection (i.e., variable selection) is to eliminate redundant or *'non-informative'* attributes from the model, especially in applications with large numbers of attributes because non-informative attributes can decrease the model's effectiveness and introduce uncertainty to the model [298]. Furthermore, the amount of data that are required to obtain reliable results is exponentially proportional to the number of features, in what is as known as *'the curse of dimensionality '* [299]. Generally, models are more easily interpreted when any unnecessary variables are eliminated, particularly when the variables are not associated with the target parameter [300]. Feature selection is also important in situations where collinearity exists between predictors to avoid increasing the complexity of the model [298]. Another reason to remove redundant variables is the cost of data collection, including the monetary cost, time required to gather, preprocessing the data, and training the model, as well as the potential environmental impacts of some of the data collection techniques.

Kuhn and Johnson [298] note that feature selection methods can be categorised as either supervised or unsupervised. Supervised techniques consider the target to evaluate the importance of the predictors, while unsupervised methods ignore the target during the selection process of informative attributes. Moreover, there are many subcategories under each method and choosing the suitable technique is determined based on two factors: first, the variable data types (e.g., numerical, and categorical) of both the targets and predictors; and second, the ML algorithm used to make prediction using the proposed subset of features. In the current study, the prediction problem is supervised learning based on numerical inputs and outputs. In addition, different ML models will be employed, and their performance will be evaluated and compared. Hence, several supervised feature selection methods will be utilised.

4.2.2 Selected Models

This section aims to describe the ML models that are used in forecasting CO_2 concentration and energy consumption to establish a general understanding of each model. The justification for selecting each model will also be discussed and the underlying algorithm will be described. All of the models were implemented using several Python libraries, mainly *scikit-learn*, keras, and NumPy. Further details related to the configuration and selection of internal model parameters will be provided in the next chapter.

4.2.2.1 Random Forest

RF is a type of ensemble learning, which was developed by L. Breiman in 2001 [301]. Biau and Scornet [302] state that RF has been tremendously successful as a general-purpose algorithm. RF have been applied extensively for prediction purposes in various fields, such solar power forecasting [303], wind power forecasting [304], building energy optimisation [305], and particulate matter concentration in the atmosphere [306]. Essentially, ensemble learners combine individual ML models, which are denoted as weak learners, to form a model with higher predictive performance. These weak learners, mostly decision trees (DTs), can be regression or classification models depending on the prediction task. DTs in general have high variance, which means that they are highly sensitive to fluctuation in the data set. Models with high variance tend to accommodate individual data points rather than understanding the general trend of the entire data set, which leads to the phenomenon of overfitting. In overfitting, high predictive performance during model training drastically diminishes when the unseen data points are introduced to the model [307]. The randomness in RF originated from two aspects: random splitting of features into smaller subsets of features, and then developing individual models using different subsets. This process of allocating different features to different DTs has the benefit of preventing certain features from dominating the prediction process. The second source of randomness introduced to RF is data sampling using, for instance, the bootstrap resampling method.

RF comprises several prediction models, which are also known as base estimators. DT was the base model that was used throughout this thesis. As shown in Figure 4.2, the learning process starts by resampling the original data by randomly selecting data points from the original data set with replacement using bootstrap resampling method, such that each bootstrap sample has the same number of data points as the original data set. This step is optional but highly recommended to avoid over-fitting. Then, each DT will be assigned a data set comprised of a subset of features. Recursively, each DT split the data at each node using the attribute that decreases impurity the most at the child node. In DT for regression, this condition is referred to as variance reduction, which is the criterion used to measure the quality of the split. This process terminates when arriving at leave nodes (i.e., nodes at the end of the tree structure containing the target values) with certain degrees of impurity. Finally, the meta model, RF in this case, takes the average of all of the predictions made by the individual trees. The predictive performance of the meta model can be measured using mean absolute error, mean squared error, R^2, and many others.

4.2.2.2 Artificial Neural Networks

ANNs are computational intelligence models that are inspired by the mechanism of the human brain. ANNs demonstrate the ability to model non-linear problems and identify the complex relationship between the inputs of a model [308]. The

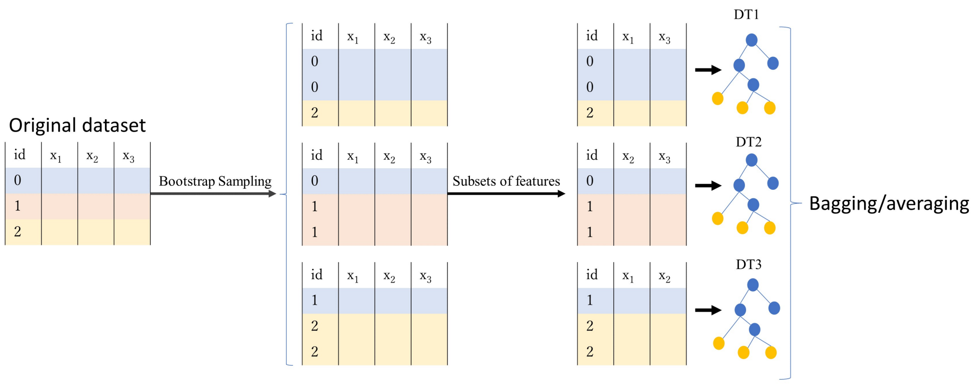


Figure 4.2: Schematic diagram of a generic RF model

first ANN model was developed by Frank Rosenblatt based on his work on the perceptron algorithm [309], which forms the basis of ANN models. The development of ANNs has experienced cycles of failure and success, with several attempts to the revive the field circa 1980s, such as the discovery of backpropagation algorithm that ANNs use for error minimisation [310]. In general, the different ANN models vary based on the connection patterns between the model components, the error minimisation process, and the type of activation function used [310]. Based on the reviewed papers in chapter 2, the multilayer perceptron is the most adopted model, a feed-forward backpropagation ANN-based architecture, hence, this model has been used throughout this thesis.

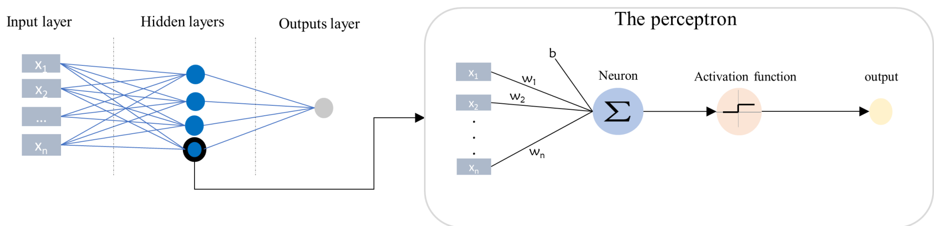


Figure 4.3: Schematic diagram of a generic ANN model

The basic structure of a generic feed-forward ANN model has three layers: an input layer, hidden layers, and an output layer. The input layer consists of nodes, each of which represents a variable to be used to train the model. The hidden layers can be a single layer or multiple layers, depending on the model architecture. The nodes within this layer are known as neurons, which carry out the calculation of a function relevant to the problem and then return some values. These values present the prediction results, which are held in the output layer. The fundamentals of ANN can be understood by looking at the perceptron algorithm, which is the building block of ANN. As shown in Figure 4.3, a perceptron is a combination of input nodes that are connected via edges to a central node (i.e., neuron), which solves a mathematical function (e.g., a linear regression equation). Then, the perceptron calls a special function, known as an activation function that returns the result of the prediction.

An example of a numerical prediction problem using ANN will be discussed for clarification purposes. The algorithm starts with initial weights, randomly generated weights, for each of the input attributes. At the neuron, a summation function computes the sum of all inputs multiplied by their weights. Then, to generate the prediction, the result is transferred through an activation function, of which there are several types (e.g., sigmoid, ReLU, and Tanh). Finally, because the initial weights are randomly generated, the prediction error is most likely to be high; hence, the algorithm iteratively updates the weights to minimise the different between the true and predicted values.

This study first developed and compared the performance of RF and ANN models in predicting CO_2 levels using various input parameters. Next, an ANN model was created to simultaneously predict both energy consumption and indoor CO_2 levels based on specific input variables. The decision to exclusively use ANN in this second prediction task is based on two observations. Firstly, in the prediction of CO_2 levels in this study, ANN marginally outperformed RF. Secondly, previous studies have shown that ANN is more popular and slightly performed better than RF in predicting energy consumption [251, 311]. Hence, it was reasonable to test ANN for the second task. However, it's important to note that future studies should not rule out testing both ANN and RF in multi-output prediction tasks, and further investigation into the capabilities of RF is warranted.

4.3 Genetic Algorithms-based Optimisation

Genetic Algorithms (GA) are a type of optimisation algorithm, and are based on the principles of natural selection and genetics. They were first introduced by John Holland as a computational model for simulating evolution [312]. GAs are used to find the best solution from a large set of potential solutions to a complex problem. This is achieved through a process of generating, evaluating, and selecting candidate solutions, which are then recombined and mutated to generate a new generation of solutions. This process continues until a satisfactory solution is found or a termination criterion is met.

In the context of a building's energy consumption, GA have been applied to optimise various aspects of a building's performance, including HVAC systems, lighting systems, building envelope characteristics, and retrofitting strategies [313, 314]. GA have been chosen as the optimisation strategy in the current research for several reasons. First, GA have proven to be effective in optimising building energy consumption by improving various building performance aspects, as is evident from the literature review in Chapter 2. These methods have been successfully applied in both commercial and residential buildings, resulting in significant energy savings. Second, GA are well-suited for complex optimisation problems, where the relationship between the inputs and outputs is not straightforward. The ventilation system, in general, has multiple design parameters that interact in non-linear ways. This makes it difficult to find an optimal solution using traditional optimisation methods. Genetic algorithms can effectively handle these types of problems by using a population-based approach that allows for exploration of the design space.

GA are a robust optimisation method that can find a satisfactory solution even when the problem is subject to uncertainty or when the solution space is large. In the context of the ventilation system, there may be uncertainty in the inputs, such as air temperature, occupancy level, humidity, and pressure, which can impact the energy use of the system. By incorporating these uncertainties into the genetic algorithm optimisation process, it is possible to obtain a solution that is robust to variations in the inputs. Finally, GA are flexible and can easily be adapted to different optimisation objectives. In the current case study, the objective is to minimise energy use. However, other objectives (e.g., minimising cost and enhancing indoor comfort) could also be incorporated into the optimisation process. This flexibility makes genetic algorithms a versatile optimisation method that can be applied to a wide range of problems.

A flowchart illustration of the steps involved in the optimisation process is given in Figure 4.4. The process of GA typically starts with a random initial population of solutions. These solutions are then evaluated and assigned a fitness score, which is a measure of how well the solution fits the problem requirements. If the current generation is less than the total number of generations, then the selection process begins. The purpose of the selection process is to identify the best solutions in the population. This is usually done through the use of selection operators, such as tournament selection, roulette wheel selection, or stochastic universal sampling. After selection, the next step is variation, where new solutions are generated (i.e. offspring) by combining the information of the best solutions. This can be done through various methods, such as crossover and mutation. The outcome of the variation process is then evaluated and assigned a fitness score. A survivor step is then performed to determine which solutions will continue to the next generation, based on their fitness scores. This process is repeated until the current generation is equal to the maximum number of generations. The final result is the best solution that has been identified by the algorithm.

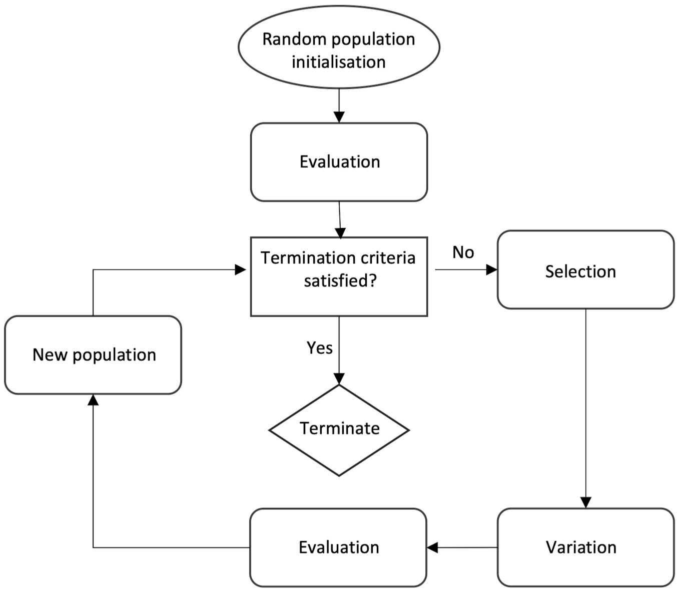


Figure 4.4: Flowchart of the GA process showing the steps involved in generating and selecting optimal solutions.

Problem formulation is an important step in using GA to find the optimal solution. To formulate a problem for the GA process, the following steps are involved:

- Identify problem parameters: this step identifies the variables that define the problem.
- Choose design parameters: from the identified problem parameters, select the variables that will be optimised using GA.
- Identify constraints: determine any constraints on the design parameters, such as physical or mathematical limitations.

- formulate objective function(s): define one or more objective functions that will be used to evaluate the fitness of the solutions.
- Choose GA parameters: set the parameters for the GA (e.g., population size, mutation rate, and crossover rate).
- Obtain solution: run the GA and obtain the best solution.
- Reformulate and rerun: if necessary, reformulate the problem and rerun the GA to obtain a better solution.

Overall, GA are a suitable optimisation method for minimising energy use of ventilation systems thanks to their ability to handle complex problems, handle uncertainty, and be adapted to different objectives. In this study, an optimisation strategy for mechanical ventilation systems using GA was developed. The objective here was to minimise energy consumption while ensuring that indoor CO_2 levels remain within acceptable limits. This strategy employed a ML model that was trained on energy simulations to predict energy consumption and CO_2 levels, which was then used as the fitness evaluation function for the genetic algorithm. It is important to carefully formulate the problem to ensure that the GA can effectively find the optimal solution. Proper problem formulation requires a clear understanding of the problem and its requirements. The parameter settings of the GA—including important aspects such as design variables and objective functions, population size, and mutation rate—will be presented in the next chapter.

4.4 Energy Simulation Model

The integration of energy simulation in the framework is underpinned by several considerations, particularly in the context of optimising the mechanical ventilation system. Indeed, historical energy data sourced from utility meters are instrumental in understanding the real energy consumption of the studied system. However, the current study required a deeper understanding of the various control strategies, which goes beyond simple energy consumption data. The optimisation process involves testing and evaluating the impact on both energy consumption and indoor conditions of an expansive design space (i.e., the set of all possible configurations of variables and operating conditions). Therefore, collecting data through practical experimentation covering all the possible configurations would be infeasible. Moreover, energy simulation aligns closely with established industry and research best practices [240, 250, 251], where simulation models serve as an important tool for system optimisation in real-world engineering context.

A thorough understanding of the case study and a requirements elicitation process are required prior to the development of a simulation model for the case study. The elicitation process has several phases. First, interviews with the facility management team are used to explore and understand the building operation strategies and the technical systems installed in the building. Then, a walk-around audit was conducted to gather information about the existing equipment, space occupancy, and to have a general understanding of the current energy consumption requirements. The next step was to review the technical specification documents and engineering drawings to develop a detailed building model. The final step was to develop a simulation model based on the gathered information using an energy simulation tool.

EnergyPlus is an open source, whole building energy simulation software that can be used to simulate energy consumption, heating/cooling load, ventilation, plug-in loads, and lighting[1]. The generated outputs are based on physics principles and complex mathematical models that are handled by the simulation engine. In its original form, EnergyPlus was programming-oriented software that received inputs and generated results as text files. However, it can be integrated with

[1] https://energyplus.net/

several user-friendly graphical interfaces, such as OpenStudio and DesignBuilder. EnergyPlus is platform-agnostic and runs on many different operating systems. It can also be integrated with other programming languages, such as Python, through a readily available application programming interface (API). Furthermore, EnergyPlus has many other features and capabilities, including allowing the users to set up simulation time steps, supporting interoperability with other simulation engines, providing advanced thermal balance models, providing pre-configured control strategies for HVAC and lighting systems, and allowing users to customise output reports. Hence, EnergyPlus was utilised in this research because of its multi-functionality and cross-platform support. Figure 4.5 illustrates a simplified workflow of EnergyPlus.

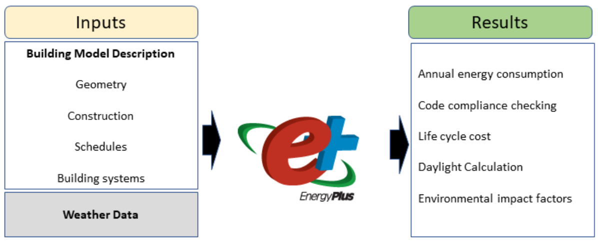


Figure 4.5: A simplified process workflow of EnergyPlus showing main inputs and outputs.

Building Model Description This represents the main input data and contains information about the buildings geometry, spaces and thermal zones, construction schedule, occupancy data, and building systems.

Weather data Weather information is an essential input to the simulation engine and has a specific file format, EnergyPlus Weather (EPW). The weather file for the case study location was obtained from Cardiff Airport weather station.

Once all of the relevant data are collected and the simulation parameters are set, the tool will carry out the simulation process and output the results with different level of details based on the number of modules involved and user preferences.

4.5 LCA Methodology

LCA is a methodology that provides accounts of environmental impacts throughout the whole life cycle of products, services, or built assets. ISO 14040 "Environmental management — Life cycle assessment — Principles and framework" [315] and ISO 14044 "Environmental management — Life cycle assessment — Requirements and guidelines" [316] have been established to standardise the processes and requirements for conducting LCA. This method has four iterative steps: i) goal and scope, ii) LCI, iii) LCIA, and iv) interpretation. In the European context, a series of relevant standards have been developed for the construction industry that are in conformity with the ISO standards for LCA, as follows: i) EN 15978 [317] for the assessment of environmental performance of buildings, and ii) EN 15804 [318] for the environmental assessment of construction products.

This section describes the design of the LCA study following the guidelines and requirements of these standards, which is broken down into four main phases, as illustrated in Figure 4.6. These are often referred to as the study design parameters, which provide quantitative and qualitative descriptions for the LCA study. It starts by describing the goal and scope of the LCA study, including the development of different scenarios and the the underlying assumptions that are applied to the study. Following this is a process to establish the inventory data and the main data sources. The impact assessment method that is adopted to evaluate the inventory data is then specified and the process ends with the interpretation phase.

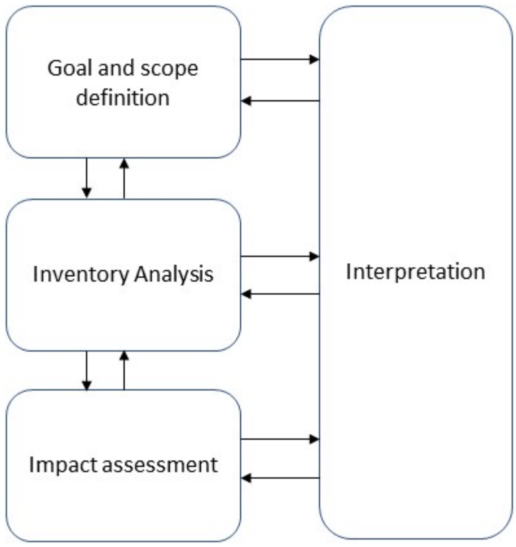


Figure 4.6: Life Cycle Assessment framework according to ISO 14040

4.5.1 Goal and Scope

ISO 14040 requires the goal of an LCA to clearly state the intended application, the reason for conducting the study, the target audience, and whether the findings will be made available to the public. In the context of building design and construction, an LCA can pursue various goals; for example, one goal is to compare different design alternatives to identify the one with the least environmental footprint, which can involve quantifying and assessing the environmental impacts of each alternative, including energy and resource use. Another goal is to eliminate building materials with the highest environmental impacts by comparing and analysing their upstream and downstream impacts (e.g., extraction, production, transportation, use, and disposal). Additionally, comparing different construction methods based on their environmental impacts is another goal of LCA studies. The literature review in Chapter 2 has provided insight into relevant themes and issues that can guide the selection of appropriate LCA goals; hence, these themes serve as good examples of LCA goals.

Following the LCA requirements and guidelines, the goal of this study is to compare different building operation strategies—namely a baseline operation strategy using static control and an optimised demand-controlled operation strategy—on the basis of environmental impact. The decision to choose one strategy over another is often difficult and complex, and several aspects should be considered. Therefore, this study intends to improve the facility manager's awareness of the environmental footprint of the building's operation strategies, and inform the effective operation and management strategies of buildings.

Functional Unit

Another important aspect at this stage is to define the functional unit of the product or system being investigated. This serves to define the function(s) of the system, and allow other practitioners to make sound and fair comparisons. However, this can only happen if everyone refers to the same function. An example of a functional unit is a whole building, which is a typical functional unit in the building domain. However, when assessing specific systems within a building, functional units other than the whole building must be defined. For instance, if the goal of the LCA study is to focus on a particular system, such as the mechanical ventilation system, then a functional unit that corresponds to the performance of that system must be used.

This study considers the mechanical ventilation system and its function, which is to supply fresh air to the occupants and remove pollutants from the indoor spaces. The performance characteristics of the identified function can be defined by the functional unit, whereby all of the system's inputs and outputs are related to a reference flow, which is required to fulfil the intended function. The functional unit that is employed in this study provides fresh air to an indoor space in an educational building with an area of 323 m^2 and a design capacity of 200 people at a recommended ventilation rate of 3 $L.s^-1$ per person.

System Boundary

Defining the system boundary is a crucial step in conducting an LCA study because it establishes the scope and extent of the assessment. The system boundary outlines the processes and activities that are included or excluded from the product system being analysed at each stage of its life cycle. For example, in the case of a whole building, the system boundary would encompass all of the life-cycle stages of the building, from design, through construction, use phase, and to decommissioning. Within each stage, the unit processes involved must also be specified to ensure consistency and comparability. For instance, during the design stage, the extraction of raw materials, manufacturing of construction products, and transportation to the building site are included within the system boundary. However, the inclusion or exclusion of these activities is subject to the goal and scope of the study, and should be clearly defined to avoid any ambiguity.

The product system under investigation is a mechanical ventilation system, with a focus on evaluating the energy consumption during the operation phase. The scope of this study is limited to a one-month time frame, during which no anticipated maintenance or replacement of components is expected. The production of raw materials, manufacturing the system's components, and replacing individual components over the lifetime of the building were not considered because this study is centred on an existing system, without modification or addition to the technical systems. The system boundary is thus defined to include the energy use associated with the mechanical ventilation system during the use phase only.

Although sensors were employed in the second scenario to monitor indoor conditions and optimise the operation of the mechanical ventilation system, they were not included in the system boundary because of the unavailability of the LCI data for the sensors. Moreover, the sensors were manufactured by different manufacturers and the materials used in their production were not obtainable, which made it impractical to establish the LCI of the sensors. However, the exclusion of the sensors from the system boundary may have implications for the accuracy and completeness of the LCA results, and may overestimate the potential environmental benefits of the optimised scenario. Additionally, the production and disposal of the sensors may have environmental impacts that were not accounted for in the study. Therefore, the results of the comparative assessment should be interpreted with caution, and future studies should consider including the life-cycle impacts of the sensors and other components that were excluded from the present study to provide a more comprehensive assessment of the environmental performance of the mechanical ventilation system.

Given that the energy consumption of the ventilation system is the primary focus of the LCA study, it is crucial to understand the electricity mix of the region or country in which the case study is located. Therefore, the electricity generation mix in Wales was analysed to establish a data inventory for the LCA calculation. This analysis is based on information presented in a report produced by Regen, a not-for-profit centre of energy transformation[2]. The report provides key statistics and analyses electricity generation from fossil fuels and renewables in Wales, as presented in Table 4.1. The analysis reveals that fossil fuels generate approximately 67% of the total electricity generated in Wales. Gas is the main source of electricity in Wales and contributes to a large portion of the total electricity generated. Diesel and waste also contribute to the total electricity generation, although their contribution is relatively small. Notably, all coal-fired plants have been decommissioned since 2020. The analysis also highlights that renewables contribute to approximately 33% of the total electricity generated in Wales. The renewable sources include onshore wind, offshore wind, solar, hydropower, and other renewables (e.g., anaerobic digestion, biomass, and landfill gas).

[2]https://www.regen.co.uk/publications/energy-generation-and-use-in-wales/

Table 4.1: Key statistics of electricity generation in Wales

Technologies	Estimated electricity generation (GWh)	Electricity generation(%)
Total	**23,102**	**100**
Fossil fuels	**15429**	**66.79**
Gas	14,425	93.49
Diesel	868	5.63
From Waste	136	0.88
Coal	0	0
Renewables	**7,673**	**33.21**
Onshore wind	3,070	40.01
Offshore wind	2,226	29.01
Solar PV	962	12.54
Hydropower	365	4.76
Other renewables	1,050	13.68

Based on this analysis, the unit processes selected for the LCA study include electricity production from gas-fired power plants, onshore wind, offshore wind, solar, and hydropower. These sources were chosen because they contribute the highest shares of electricity production in Wales, as reported in the Regen report. For example, gas-fired power plants are the primary source of electricity in Wales, generating approximately 63% of the total electricity in 2020. Onshore wind is also a significant contributor, accounting for approximately 13.29% of the total electricity generation in 2020. In contrast, sources such as diesel and other renewables (e.g., anaerobic digestion, biomass, landfill gas) had relatively small contributions to the total electricity generation in Wales. Therefore, these sources were combined into a single unit process called the "production mix". This decision was also made due to the lack of available environmental data for each of these technologies. The resulting unit processes that were included in the product system are depicted in Figure 4.7

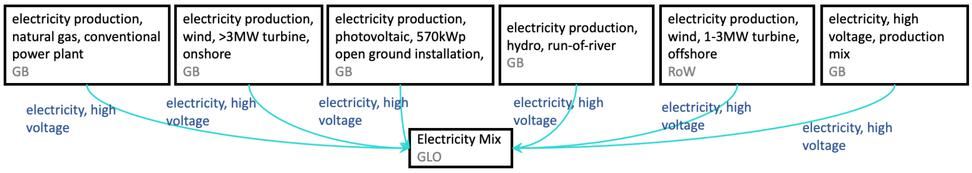


Figure 4.7: Flow-chart showing unit processes considered in the electricity mix.

4.5.2 Life Cycle Inventory

Upon defining the system boundary, the next stage is to establish the LCI of all of the unit processes that have been identified within the system boundary. The primary objective of this stage is to collect the inputs and outputs of the unit processes involved in the studied system. The data collection process can be accomplished via primary or secondary sources. The primary sources refer to data that are generated directly by the LCA practitioners, whereas the secondary sources include data from the literature, manufacturers, and LCA databases (e.g., the ecoinvent database). The output of this stage is a comprehensive data inventory that includes all of the relevant data that have been obtained from the identified unit processes. This inventory serves as the input to the subsequent LCIA stage.

The LCI of the study was established using a combination of primary and secondary data. The primary data were obtained from energy simulations and optimisation conducted in the previous sections for both the baseline and optimised scenarios. These data relate to the energy consumption of the ventilation system in the building. Meanwhile, the secondary data were collected from the Ecoinvent 3.8 database for all of the unit processes involved in the electricity production.

The LCI that was employed to produce the functional unit is given in Table 4.2, which presents a comprehensive list of all the unit processes considered. This table further details the associated Ecoinvent process of each unit process and their respective units of measurement. The data that are presented in this table form the basis for the impact assessment and the ultimate evaluation of the environmental performance of the mechanical ventilation system under study.

Table 4.2: Data sets used for unit processes

Unit process	**Unit**	**Ecoinvent process**
Electricity production (natural gas)	kWh	electricity production, natural gas, conventional power plant (GB)
Electricity production (onshore wind)	kWh	electricity production, wind, >3MW turbine, onshore (GB)
Electricity production (offshore wind)	kWh	electricity production, wind, 1-3MW turbine, offshore (RoW)
Electricity production (Hydropower)	kWh	electricity production, hydro, run-of-river (GB)
Electricity production (Solar)	kWh	electricity production, photovoltaic, 570kWp open ground installation, multi-Si (GB)
Electricity production (mix)	kWh	electricity, high voltage, production mix (GB)

4.5.3 Life Cycle Impact Assessment

During this stage of the LCA, the focus shifts from the inventory results to the assessment of the potential environmental impacts that are associated with the product system under consideration. A variety of impact assessment methods are available in the literature, which incorporate different impact categories (e.g., climate change, ecotoxicity, human toxicity, and ozone depletion, etc.). Well-known impact assessment methods include TRACI, ReCiPe, IPCC, and IMPACT 2002+. Essentially, impact assessment methods consider each environmental stressor (e.g., carbon emissions) and apply a series of steps (e.g., selection, classification, and characterisation) to create various impact categories that are relevant to resources, human health, and other areas of concern. The outputs of this stage are a list of impact categories, which can be analysed and prioritised based on the intended goal of the study.

It is important to note that the selection of impact assessment method and the impact categories depends on the goals and scope of the study, which may vary depending on the specific product system being evaluated. In this study, the goal is to compare the environmental performance of the baseline scenario and the optimised scenario for the mechanical ventilation system. Note that the scope is limited to the energy consumption during the operation phase of the system. The impact categories that were considered include:

- Climate change: This impact category is relevant because the study is focused on energy consumption, which is directly related to GHGs, which contribute to climate change.

- Human toxicity: The inclusion of the impact category in the impact assessment phase is important because of the use of hazardous chemicals in the production of certain components used in renewable technologies, including PV panels. These chemicals have the potential to cause harm to human health, which makes it necessary to consider this impact category in this study.

- Fossil fuel depletion: This impact category was considered in this case because gas is the main source of electricity generation in Wales. This is an indicator of the finite availability of natural resources.

- Metal depletion: This can be an important impact category to consider in the environmental assessment of renewable energy technologies, such as wind turbines, because they require a significant amount of metals for their production and installation.

The ReCiPe Midpoint method was chosen as the impact assessment method in this study because it covers all of these impact categories.

4.5.4 Interpretation

The final stage of the LCA is interpretation, which analyses the results in the context of the study's objectives. During this stage, valuable insights can be drawn to inform the decision-making process. In addition, recommendations can be provided to select the most suitable alternative or identify areas for improvement to minimise environmental impacts.

4.5.5 LCA Modelling Tool

Numerous tools are available for modelling LCA studies, with both open source and proprietary software options. In the present study, Brightway2, an open source software written in Python, was employed to develop the LCA model. Brightway2 can be utilised in two ways: through a Jupyter notebook or via the Activity Browser (which is a graphical user interface that employs Brightway2 in the background). Brightway2 provides users with a range of features, such as the capability to manage large-scale data, develop complex models, and interface with various databases. Brightway2 can also be integrated with other tools via an API.

4.6 Dynamic Data Collection and Preparation

To develop the prediction models, the data were collected over two months (i.e., August and October). The plan was to include September data, but technical difficulties with the monitoring system caused most sensors to go offline due to power supply issues with the gateway. This problem was not resolved until the end of the month, resulting in the unavailability of the data for September. Nonetheless, as September falls within the university's summer holiday period, the August data can be deemed to be representative of the summer period and the indoor space dynamics were still captured, albeit with some loss of training data. The data collection and pre-processing process involved multiple stages. For the sake of brevity, discussion and figures presented here only reflect the August data. Nevertheless, a similar workflow was followed for the October data.

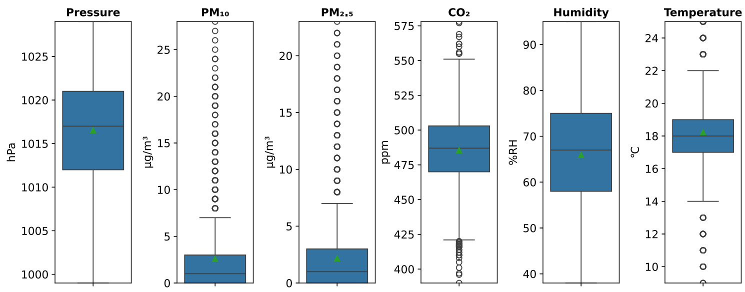
The data were collected from three sources, namely indoor remote units, on-site weather stations, and window sensors. The sensed data were transmitted to the database at varying sample rates. These differences in sampling intervals were primarily caused by the pre-configured measurement frequency of each device during the assembly and manufacturing process. Another contributing factor for the variation in sampling intervals was related to the intended application of the sensing device. For instance, window sensors reported data based on state changes (i.e., the opening and closing of windows), which typically occurred at irregular intervals. The inclusion of window sensors resulted in the data being reported on an event-driven basis, with state changes occurring sporadically, while indoor remote units and on-site weather stations captured data at a fixed frequency. Therefore, the data that were collected from these sources presented varying sampling rates, which necessitated some preprocessing (mainly resampling) to synchronise and align the data for analysis and modelling purposes.

Resampling refers to the process of modifying the frequency of time series data, and it involves two main methods: down-sampling and up-sampling. Down-sampling reduces the granularity of data, such as converting hourly counts to daily counts. This is done through statistical aggregation methods, such as mean or median, which aggregate multiple observations into a single data point. Conversely, up-sampling increases the level of granularity by generating new data points within the existing observation range using interpolation. Up-sampling is employed when a specific application requires higher-frequency data than what is available in the database. In this study, both down-sampling and up-sampling methods were utilised prior to the data consolidation step to address the varying sampling intervals of the three data sources. These techniques allowed for the synchronisation and alignment of data, and enabled the development of accurate prediction models.

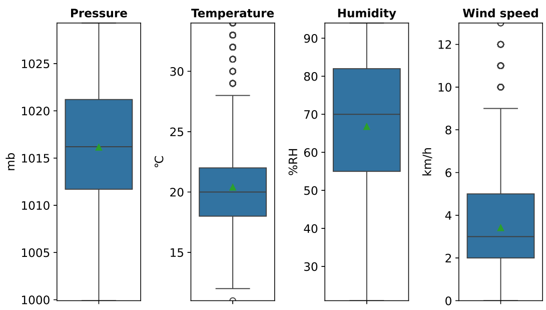
Figure 4.8-(a) illustrates the distributions of indoor parameters, including pressure, CO_2 level, humidity, temperature, PM10, and PM2.5. For pressure, CO2 level, humidity, and temperature, the average value was used for aggregation due to the near balance of both sides of the distribution. However, for PM10 and PM2.5, the median was employed instead of the mean because of the positive skewness that was caused by extreme values that shifted the mean away from the centre of the distribution.

Similar considerations were applied to the outdoor parameters, including pressure, temperature, humidity, and wind speed (Figure 4.8-(b)). In this case, the mean was used for pressure and temperature, while the median was employed for humidity and wind speed due to the apparent left and right skewness, respectively.

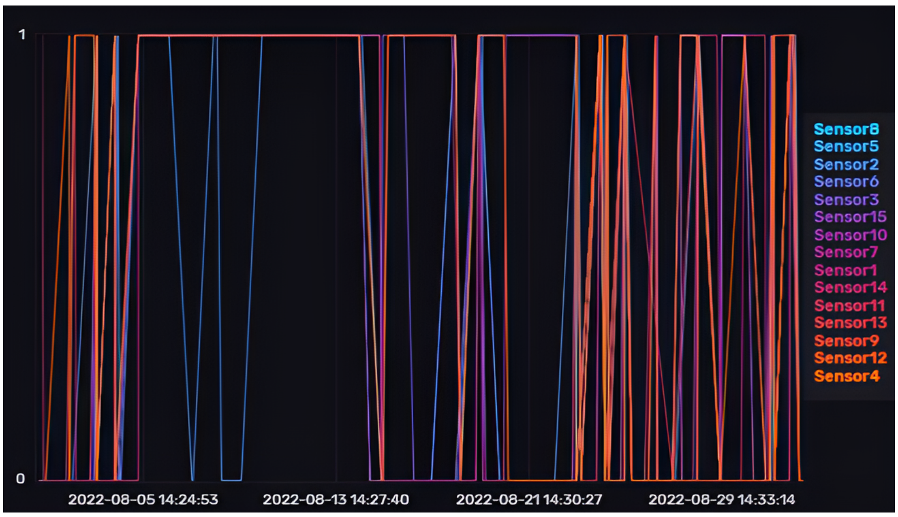
Figure 4.8-(c) depicts the binary data from the window sensor, where a value of zero represents a closed state, and a value of one denotes an open state. These open/close events are stochastic and can occur at any time. Once a state is triggered, it persists until the opposite state is activated. The time span between two distinct states may vary from minutes to days. Therefore, these data cannot be combined with other regularly sampled time-series data. To address this issue, a preprocessing step and resampling were carried out on the window sensor data. The raw data were transformed into a time series with a frequency of 1 minute, as demonstrated in Figure 4.9, by means of resampling to eliminate irregularities in the raw data.



(a) Boxplots for sensor readings for several indoor environmental parameters



(b) Boxplots for outdoor environmental conditions from the weather station



(c) Window sensors profile over the monitored period

Figure 4.8: Visualisation of the raw data prior to resampling

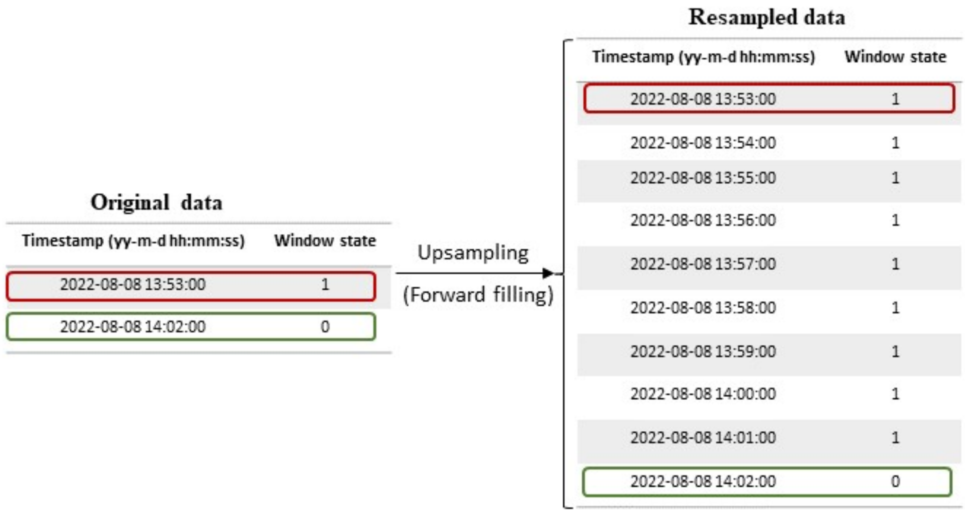


Figure 4.9: Illustration of an example of the resampling process for the window sensor data .

After resampling was conducted and all of the data sources had regular sampling intervals, the next step was to consolidate the data into a single data frame. The desired data interval was set to hourly frequency because the prediction models were designed to predict outputs on an hourly basis.

4.7 Summary

The two research questions associated with this chapter were:

Can access to dynamic data provide more accurate accounts of the environmental impacts during the operation stage?

How can machine learning and optimisation be leveraged to reduce the environmental impact of buildings?

This chapter laid the foundation for answering these two research questions, although it did not provide direct answers. Instead, it discussed the necessary steps that must be taken to leverage dynamic data for ML and optimisation models. Additionally, the theoretical background for the modelling choices was discussed, which provided insight into their underlying principles.

The chapter discussed two data-driven prediction models used in predicting CO_2 concentration and energy consumption, namely RF and ANN. The choice of RF and ANN models for CO_2 concentration and energy consumption prediction was justified by their success in various fields, and their ability to handle non-linear relationships and high-dimensional data. Additionally the use of genetic algorithms was reviewed and considered in this study due to its ability to optimise complex problems. Furthermore, the LCA model, which is a core component of the framework, was explored. This chapter also explained how the model should be developed to provide a comprehensive understanding of the studied system.

This was followed by a discussion of the necessary steps for leveraging dynamic data in the context of ML and optimisation for environmental impact assessment during the operation stage. Specifically, a thorough preprocessing methodology was implemented to transform the raw dynamic data from multiple sources into a single, consolidated data frame with a regular hourly time series format. The preprocessing steps included data resampling and aggregation, which were customised to each parameter based on its specific frequency. The resulting data frame will be used as input for various ML models and optimisation algorithms in the next chapter to predict and optimise the environmental impacts during the operation stage.

In conclusion, this chapter has highlighted the key aspects of developing a robust LCA framework, including dynamic data, ML, optimisation, and the LCA model. These aspects are essential to ensure that the LCA results are accurate, credible, and useful for making an informed decision. The next chapter will present the implementation of the framework that was developed in the current chapter and its potential implications.

Chapter 5

Outputs and Results

This chapter details the implementation of the proposed framework that was outlined in Chapter 4. The primary focus of this chapter is to address the following research questions:

Can access to dynamic data provide more accurate accounts of the environmental impacts during the operation stage?, and

How can machine learning and optimisation be leveraged to reduce the environmental impact of buildings?.

The chapter begins by exploring and contextualising the dynamic data that are incorporated into the framework. Next, a detailed discussion of the developed ML-based prediction models is presented. The optimisation strategy and associated algorithms are also described in depth. The implementation of the LCA model is then presented, with a focus on how it is integrated into the overall framework. Finally, this chapter concludes by revisiting the research questions and providing answers based on the findings presented throughout this chapter.

5.1 CO_2 Prediction Models

This section presents the results of the work that was conducted on the prediction of CO_2 levels in the Forum. Several ML models were developed to forecast CO_2

levels at two different time horizons: 1-step (1-hour ahead) and 24-step (24-hour ahead) prediction. For the 1-step prediction, two approaches were employed: the first approach utilised univariate models that solely relied on CO_2 lags to predict the CO_2 level in the next hour, and the second approach utilised multivariate prediction models that included not only CO_2 lags but also other parameters that were introduced in the previous chapter. The 24-step ahead prediction was based on the best-performing model from the first approach, which was further extended to forecast CO_2 levels for the next 24 hours. The following subsections will provide a detailed description of each approach and the results that were obtained.

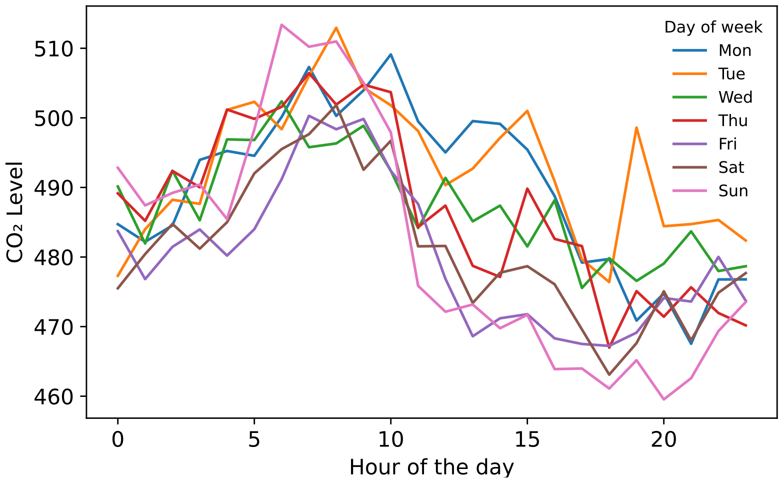
5.1.1 Description of the CO_2 Observations

Figure 5.1 gives the average CO_2 levels per hour for each day of the week over a 24-hour period in the Forum during August and October. It provides the average CO_2 levels starting from hour 0, which corresponds to the beginning of the day (i.e., midnight), and ending with hour 23 (i.e., 11 p.m.). It is important to note that there is a one-hour gap between data points for consecutive days. For example, the last recorded CO_2 level on Wednesday is at hour 23 (i.e., 23:00), and the subsequent data point is at midnight on Thursday (i.e., hour 0 or 00:00). Given this one-hour gap, it is improbable that the CO_2 level remained constant during this period.

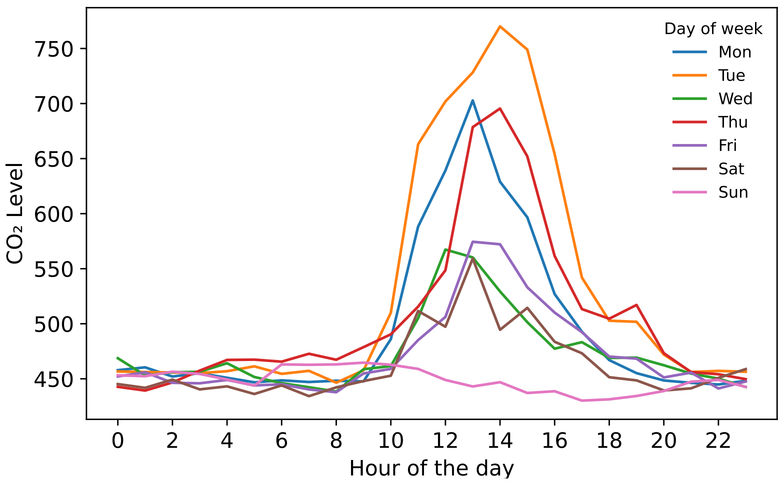
In August, the CO_2 levels fluctuate between 450 and 510 ppm. This can be attributed to the lower occupancy levels in the space during the summer holiday, with fewer students and individuals using the space to work or socialise. In addition, the relatively stable CO_2 levels in August may suggest adequate ventilation in the space. However, when taking a closer look at Figure 5.1-(a), an unexpected trend becomes apparent. Specifically, throughout the month of August, the figure shows a notable CO_2 peak occurring consistently between 6 to 7 a.m. each day, surpassing levels observed during the same hours in October. This observation is noteworthy, as it cannot be attributed to occupancy patterns. During the early hours of each day in August, the space is mostly unoccupied. There are two plausible explanations. First, there might be an issue with sensors giving incorrect readings (i.e., a sensor malfunction). However, this issue was only observed in August, with CO_2 levels appearing normal in October. Second, there is the possibility of a timestamp discrepancy in the data fed into the server from the sensors. Shifting the data by a few hours forward might yield a peak occurring in the afternoon, which is in line with our expectations.

In contrast, the October chart shows that there is a clear trend in CO_2 levels, with levels increasing from their minimum at 8 am and peaking at around 2 pm on weekdays. This trend is consistent across all weekdays, with a bell-shaped profile that reaches a peak between 550 and 800 ppm, with Tuesday exhibiting the highest levels. This trend can be attributed to the higher occupancy levels in the space during the university's autumn semester, when students and other individuals use the space more frequently for group work and socialisation. The Forum is busiest in the afternoon period, which also coincides with higher CO_2 levels. This indicates that increased occupancy causes the CO_2 levels to rise. Saturdays exhibit a trend that is similar to weekdays, with higher CO_2 levels during the afternoon when the School of Engineering uses the Forum to hold events. In contrast, Sundays have relatively stable CO_2 levels throughout the day, fluctuating between 460 and 510 ppm.

The data suggest that CO_2 levels in indoor environments are highly dependent on a number of factors, including occupancy, ventilation rates, and time of day. As evidenced by the August and October charts, CO_2 levels in the same indoor space can vary significantly between months, and even within the same month, while different days can have unique CO_2 profiles.



(a) August



(b) October

Figure 5.1: Average CO_2 levels for each day of the week in the Forum during August and October over a 24-hour time period.

This variability underscores the need for a dynamic and flexible ventilation strategy that can be adjusted in response to the changing conditions. Furthermore, a fixed schedule for ventilation may not be optimal from an energy use standpoint because it may result in over-ventilation during times of low occupancy or under-ventilation during peak occupancy times.

Predicting CO_2 levels can enable the building's managers to anticipate when ventilation rates need to be adjusted to maintain healthy indoor air quality while minimising energy use. This type of approach is crucial for reducing the environmental impact of buildings and promoting sustainable practices in the built environment.

5.1.2 Univariate One-step Prediction Models

An autocorrelation plot for one month's worth of hourly CO_2 data can provide useful information about the presence and strength of autocorrelation in the data at different lags. The autocorrelation plot can also provide important insights into the characteristics and patterns of time series data, and can be a useful tool for selecting the number of lags in a ML-based prediction model. Lag variables with correlation values close to either +1 or -1 indicate high correlation, whilst values close to 0 indicate low correlation.

The autocorrelation plot for October in Figure 5.2 shows a pattern of autocorrelation with positive correlation values for small lags, which correspond to adjacent hourly CO_2 measurements. The autocorrelation values decrease as the lags increase, which indicate a declining influence of previous observations on the current observations. There is also a repeating pattern in the autocorrelation plot at approximately 24-hour intervals, which is likely to be due to the diurnal variation in CO_2 levels in the indoor space. This suggests that the CO_2 data may exhibit daily seasonality, which could be useful for modelling and prediction.

The autocorrelation plot for August does not exhibit any clear trend or seasonality and the autocorrelation values fluctuate around zero with no discernible pattern (Figure 5.3). This suggests that the CO_2 data in August may be more random or noisy when compared to October and that there may not be any strong autocorrelation between adjacent hourly measurements.

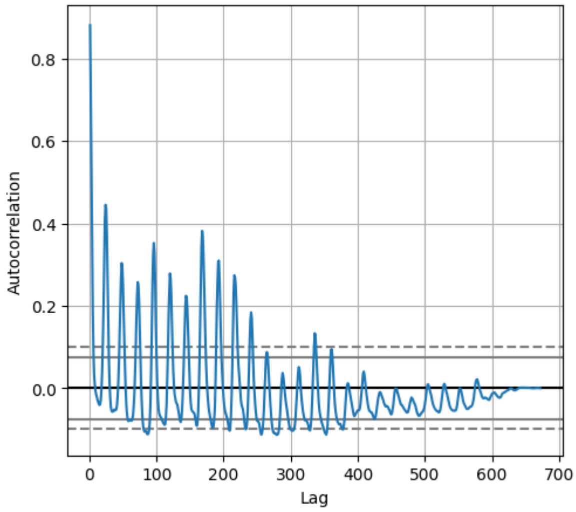


Figure 5.2: Autocorrelation plots between hourly CO_2 observations and different lag variables in October.

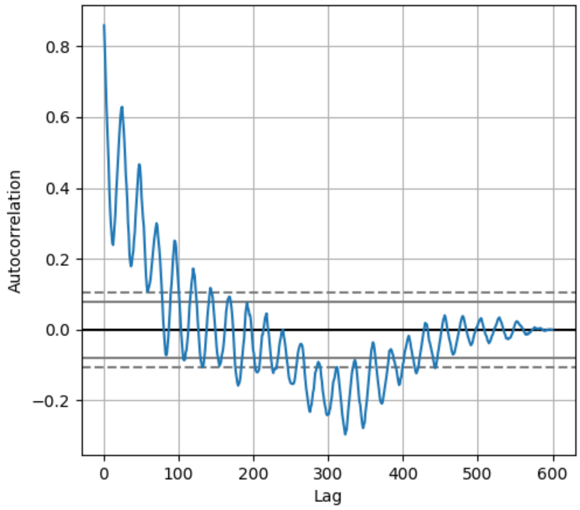


Figure 5.3: Autocorrelation plots between hourly CO_2 observations and different lag variables in August.

In terms of selecting the number of lags for a ML-based CO_2 prediction model, the autocorrelation plot for October may be more informative because it shows a clear pattern of autocorrelation that can be exploited to model the relationship between past and future CO_2 levels. In contrast, the autocorrelation plot for August does not provide any clear information about the relationship between past and future CO_2 levels. This may make it more challenging to develop an accurate prediction model. Hence, univariate ML models will be developed to predict future CO_2 levels for the month of October only. The specific number of lags to be included will be determined based on the performance the ML model.

Figure 5.4 shows a positive correlation for the first four lags, after which the correlation approaches zero. Based on this observation, it has been decided to consider only the first four lags for the ML models that were developed to predict CO_2 levels because the plot indicates that the first four lags are the most relevant for capturing the relationship between past and future CO_2 levels.

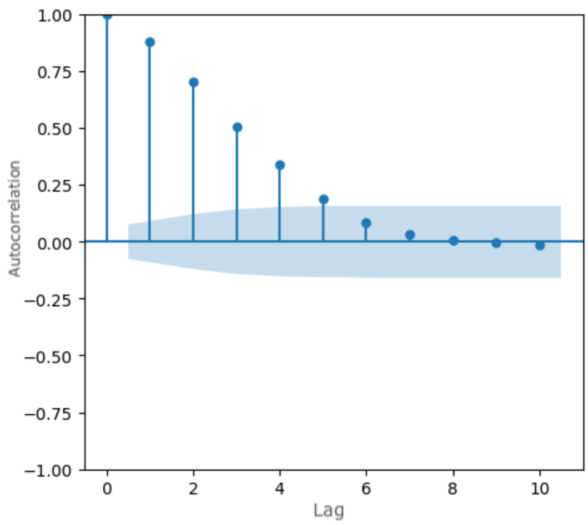


Figure 5.4: Autocorrelation plots between hourly CO_2 observations and the first 10 lags in October.

To determine the optimal number of lags to include in the univariate ML model to predict CO_2 levels in October, a range of lag values will be experimented with and the model's performance will be evaluated using appropriate metrics. The number of lags will be selected based on the best-performing model, which will provide the most accurate and reliable predictions of CO_2 levels in the Forum.

Table 5.1 presents the results of the RF models that were developed to predict CO_2 levels using October data. The table includes an ID column to identify each model configuration, a column specifying the number of estimators used in the RF model, a column indicating the number of lags considered, and a column showing the model's performance based on the Root Mean Squared Error (RMSE) metric, which measures the difference between predicted and actual CO_2 levels.

Table 5.1: Performance of univariate RF models for CO_2 prediction

Model	# **Estimators**	# **Lags**	**Error**(RMSE)
RF-01	200	1	39.920
RF-02	200	2	30.195
RF-03	200	3	22.830
RF-04	200	4	21.317
RF-05	400	1	39.920
RF-06	400	2	30.195
RF-07	400	3	22.830
RF-08	400	4	22.475
RF-09	600	1	39.710
RF-010	600	2	30.371
RF-011	600	3	23.777
RF-012	600	4	21.235
RF-13	800	1	39.448
RF-14	800	2	29.505
RF-15	800	3	24.241
RF-16	800	4	21.738
RF-17	1000	1	39.737
RF-18	1000	2	29.336
RF-19	1000	3	23.679
RF-20	1000	4	21.609

To determine the best-performing RF model, a number of configurations were tested with different numbers of estimators and lags. The performance of each individual model was evaluated using the RMSE metric. The results show that the best-performing RF model was RF-012, which utilised 600 estimators and considered four lags. This model achieved an RMSE of 21.235, which indicates that there is a high level of accuracy and reliability in predicting CO_2 levels in the space.

Table 5.2 presents the performance of different configurations of ANN models when used to predict the CO_2 level based on the historical data of lag variables. Each configuration is represented by a unique ID and the table includes information on the optimiser used (Adam and SGD), the number of neurons in the hidden layer, the number of lags used, and the performance of each individual model based on the RMSE metric. The results of Table 5.2 suggest that the choice of optimiser and the number of neurons in the hidden layer can have a significant impact on the performance of the ANN models. The best-performing model was ANN-05, which used the Adam optimiser, 10 neurons in the hidden layer, and one lag, and achieved an RMSE of 18.340.

It is worth noting here that the performance of the ANN models in Table 5.2 is generally better than that of the RF models in Table 5.1. The best ANN model achieved an RMSE that is approximately 13% lower than that of the best RF model. This suggests that, in this specific case, ANN models may be a more effective choice for predicting CO_2 levels based on the current data.

Table 5.2: Performance of univariate ANN models for CO_2 prediction

Model	**Optimiser**	# **Neurons**	# **Lags**	**Error**(RMSE)
ANN-01	Adam	5	1	168.514
ANN-02	Adam	5	2	155.611
ANN-03	Adam	5	3	145.42
ANN-04	Adam	5	4	148.936
ANN-05	Adam	10	1	18.340
ANN-06	Adam	10	2	20.507
ANN-07	Adam	10	3	19.869
ANN-08	Adam	10	4	21.653
ANN-09	Adam	15	1	18.401
ANN-10	Adam	15	2	18.434
ANN-11	Adam	15	3	19.395
ANN-12	Adam	15	4	20.976
ANN-13	Adam	20	1	18.611
ANN-14	Adam	20	2	19.096
ANN-15	Adam	20	3	18.759
ANN-16	Adam	20	4	20.343
ANN-17	SGD	5	1	35.376
ANN-18	SGD	5	2	34.589
ANN-19	SGD	5	3	34.073
ANN-20	SGD	5	4	34.536
ANN-21	SGD	10	1	31.320
ANN-22	SGD	10	2	31.827
ANN-23	SGD	10	3	30.193
ANN-24	SGD	10	4	30.356
ANN-25	SGD	15	1	30.125
ANN-26	SGD	15	2	29.434
ANN-27	SGD	15	3	30.925
ANN-28	SGD	15	4	30.279
ANN-29	SGD	20	1	31.073
ANN-30	SGD	20	2	31.459
ANN-31	SGD	20	3	30.543
ANN-32	SGD	20	4	31.259

5.1.3 Multivariate One-step Prediction models

Univariate models use only historical CO_2 levels as inputs to predict the future CO_2 level. In contrast, multivariate prediction models incorporate additional variables (e.g., outdoor conditions, indoor conditions, and time of day) to make more accurate predictions. These models take into account the complex interactions between the factors that can influence CO_2 levels, which can improve the accuracy of the predictions. Therefore, a better understanding of the impact of different variables on CO_2 levels can be achieved by using multivariate models.

RF Models

This section will describe the multivariate models that were developed to predict the CO_2 levels in the next hour. Accounting for airflow dynamics is important in understanding and predicting indoor air quality. However, integrating airflow as a parameter in the ML model is challenging in terms of data acquisition and modelling. Unlike measurable attributes, such as window status, outdoor temperature and wind speed, airflow is a multidimensional phenomenon that involves fluid dynamics, and can be influenced by several factors, including room layout, window size and type, and building location and orientation. Hence, opting for window status and measurable outdoor parameters rather than attempting to directly incorporate complex airflow dynamics is a practical decision. Furthermore, these parameters serve as a proxy for assessing the impact of natural ventilation on indoor conditions. Furthermore, it is imperative to acknowledge that the binary reading for window status, presents a limitation in capturing the extent of window openness. The inability to differentiate between full and partial window openings is a notable issue in this parameter.

For the month of August, the models include four CO_2 lags and several parameters, such as outdoor conditions (temperature, wind speed, humidity, and pressure), window status (open/closed), indoor conditions (temperature, humidity, pressure, PM-10, and PM-2.5), time of day (24 hours), and day of the week. The inclusion of window status when predicting indoor CO_2 concentrations is of particular interest since it allows for an assessment of the usefulness of this attribute in predicting CO_2 levels in the Forum. In addition, the consideration of outdoor parameters in the models is crucial because it can impact indoor conditions, including CO_2 concentrations. The inclusion of these parameters in the models provides insights into the key factors affecting indoor CO_2 levels, which can inform the design and implementation of effective ventilation strategies.

Identifying the informative parameters is crucial because it directly affects the accuracy of any prediction model. RF incorporates a built-in feature importance tool, which can effectively identify the most influential parameters for a given model. The feature importance analysis of the RF model indicated that the CO_2 level at the previous hour and humidity in the Forum were the two most important predictors, with importance values of 0.33 and 0.29, respectively (Figure 5.5). In general, higher humidity levels tend to be correlated with higher CO_2 levels because both humidity and CO_2 levels are affected by occupancy. In this particular context, it is worth highlighting that the average indoor humidity closely mirrored the outdoor humidity levels, both averaging approximately 70%. Hence, the high indoor humidity levels cannot be solely attributed to space occupancy. Furthermore, in RF models, feature importance is a measure of how important each feature is in predicting the target variable. It means that the variable is very informative and can be used to build an accurate model. However, feature importance does not explain why a feature is important or provide an explanation of the underlying causal relationships.

The remaining predictors had negligible importance values of less than 0.09. Interestingly, while external wind speed and window status are recognised factors in influencing indoor air conditions, their limited impact in this particular analysis can be attributed to the specific conditions observed in August. First, the low

CO_2 levels recorded during that period, which aligned closely with outdoor levels, suggest that it would be improbable for external factors to have a substantial impact on indoor CO_2 levels. Additionally, the average wind speed in August remained consistently below 4 km/h. This relatively low wind speed may have reduced airflow through open windows. Therefore, in this context, the interplay between the observed CO_2 levels and the low wind spend may have influenced the informativeness of external wind speed and window status in predicting indoor CO_2 concentrations. These findings were used to refine the model and improve its accuracy by focusing on the most important parameters and potentially eliminating the less important parameters.

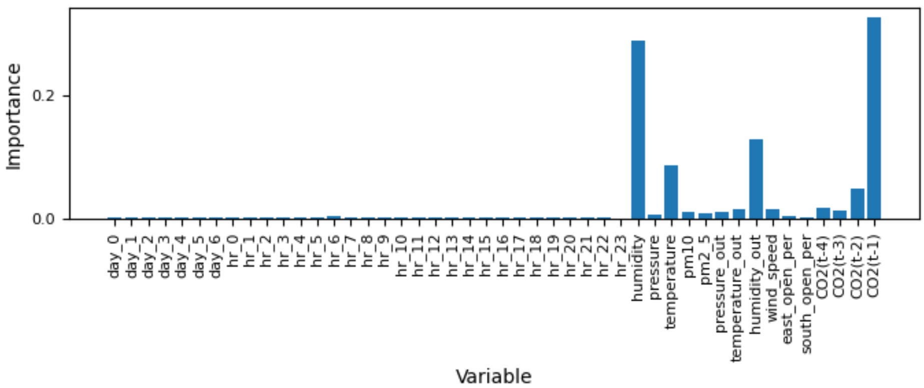


Figure 5.5: Feature importance analysis for the RF model in August

In the multivariate RF model for October, the inclusion of parameters was limited to four CO_2 lags, indoor conditions (temperature, humidity, pressure, PM-10, PM-2.5), time of day (24 hours), and day of the week. Notably, window status was excluded from the model because it was found to be irrelevant due to a lack of opening and closing events during this month. As a result, the indoor air was mainly conditioned by mechanical ventilation and outside conditions had little influence. The feature importance analysis shows that the most influential parameters are the CO_2 level of the previous hour and pressure, with importance values of 0.8 and 0.1, respectively (Figure 5.6). However, the other parameters had negligible importance values (i.e., close to 0). Interestingly, the time of day and day of the week were found to be uninformative, which is unexpected given the CO_2 peaks that were observed during the afternoon on weekdays. Further investigation may be necessary to understand the discrepancy between the results obtained in October and the findings from the August models. The limited data set that was used for this study, which only includes data from October, might have contributed to this discrepancy.

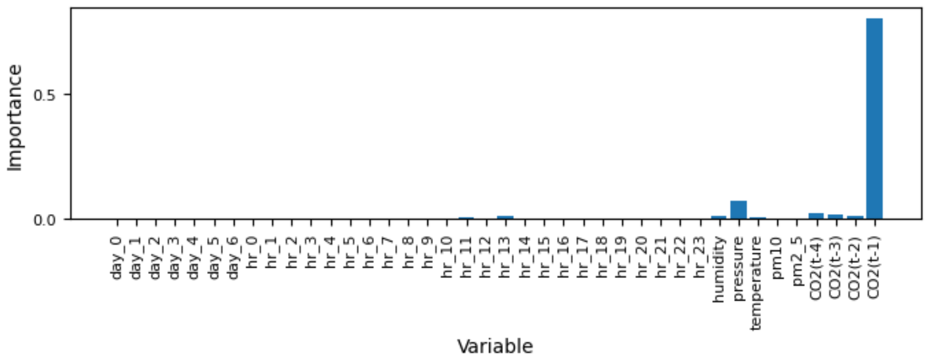


Figure 5.6: Feature importance analysis for the RF model in October

Figure 5.7 gives the actual and predicted CO_2 levels for the month of August. The x-axis of the graph represents the sample hours, while the y-axis represents the CO_2 values. The original dataset was partitioned into two distinct subsets: a training set and a testing set. This division facilitated the evaluation of how effectively the model generalises to unseen data. Initially, the model was trained on the training set. After training, the model was used to make predictions on the testing set to evaluate the performance of the model. Moreover, the sample hours used for evaluation were taken directly from the testing set. They were chosen to be consecutive hours rather than selected randomly. This sampling approach was employed to align with the nature of time series data, sequence is important. This graph indicates that the predicted CO_2 levels generally follow the pattern of the actual CO_2 levels, with some minor variations. The RMSE for this model is 18.341, which indicates that there is a relatively low prediction error. Similarly, Figure 5.8 shows the actual and predicted CO_2 levels for the month of October. The graph indicates minor variations between the actual and predicted values. The RMSE for this model is 19.523, which indicates that there is a slightly higher prediction error when compared to the August model.

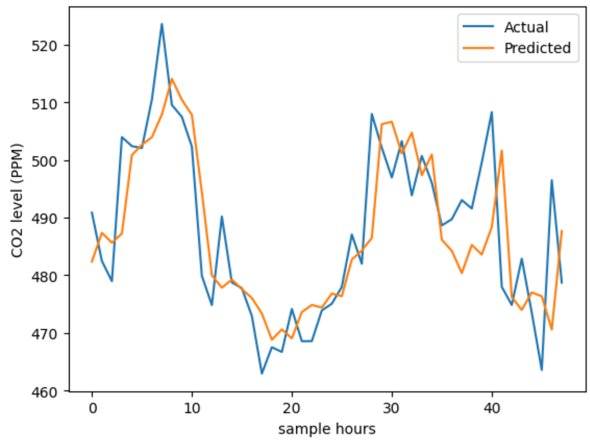


Figure 5.7: Sample hours prediction using the RF model during August

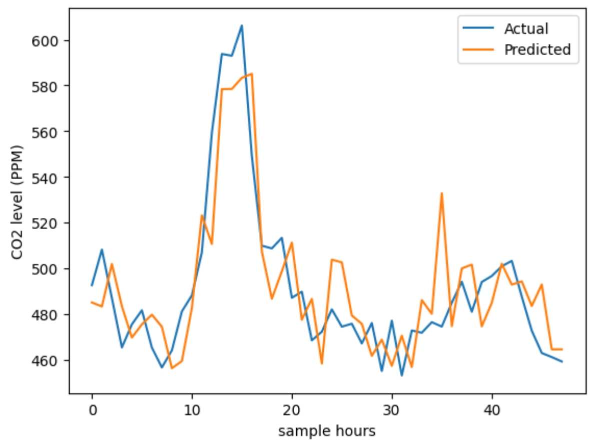


Figure 5.8: Sample hours prediction using the RF model during October

The dissimilarities in the underlying patterns of the data for the two months or the use of distinct sets of variables in the modelling process may explain the discrepancy in the performance between these two models. However, despite the slightly higher prediction error in October, both models show promising results in predicting CO_2 levels using multivariate RF models.

In general, ML models, particularly within this context, function by discerning patterns from historical data. The historical data encompasses a series of instances with different attributes, including hour of the day, day of the week, CO_2 lags, among others. In this study, the approach began with a simplified model that only relies on the previous hour's (t-1) CO_2 level to predict the CO_2 in the next hour (t). The main objective of this model is to learn from (t-1) CO_2 levels to make accurate predictions regarding CO_2 levels at (t). Therefore, this process should not be considered as a mere time shifted version of the actual data. Furthermore, initiating the modelling process with a simple model and a minimal set of attributes is a well-founded practice in ML. This is often followed by adding complexity (i.e., considering more attributes) until a satisfactory level of performance is achieved. Following the initial examination of CO_2 lags (e.g., CO_2 at t-1), additional parameters can be examined. For example, it is reasonable to anticipate that the CO_2 levels at specific times on specific days may exhibit variation (e.g., hour of the day, and day of the week). However, it is important to acknowledge that with a limited number of data points, particularly when factoring in a considerable number of input attributes, data sparsity becomes a critical issue. This phenomenon is often referred to as the "curse of dimensionality", which was mentioned in the feature selection section in Chapter 4. For example, in the dataset used, there were only 4 CO_2 readings corresponding to 2 p.m. on Mondays for each month. While this data constraint is significant, it should not discourage the exploration of these additional attributes. In the end, it is worth noting that while these attributes may not have demonstrated significant influence due to the limitations of the available data, their importance and potential relevance in other contexts should not be prematurely dismissed.

ANN Models

In the development of multivariate ANN models for predicting indoor CO_2 levels in August and October, a correlation-based approach was used to select the models' parameters. In addition, a correlation matrix was used to identify parameters that exhibit high correlation with the target variable, CO_2. The August correlation matrix showed that indoor humidity and temperature have strong positive and negative correlations with CO_2, with correlation coefficients of 0.77 and -0.65, respectively (as shown in Figure 5.9). In addition, outdoor humidity and temperature were found to have high positive and negative correlations with CO_2, with correlation coefficients of 0.797 and -0.725, respectively (see Figure 5.10). Furthermore, the correlation coefficient between window status and CO2 levels was found to be 0.184, indicating a relatively low correlation.

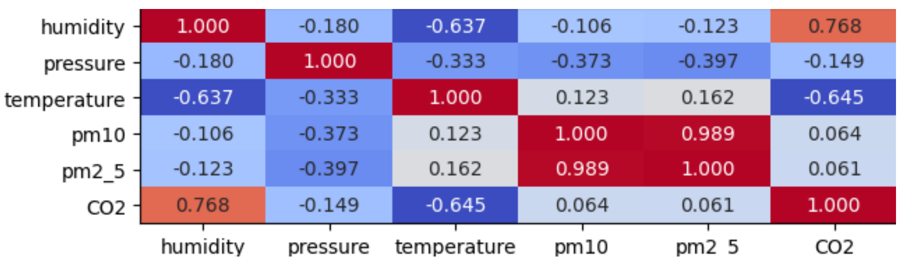


Figure 5.9: Correlation matrix of the indoor parameters in August

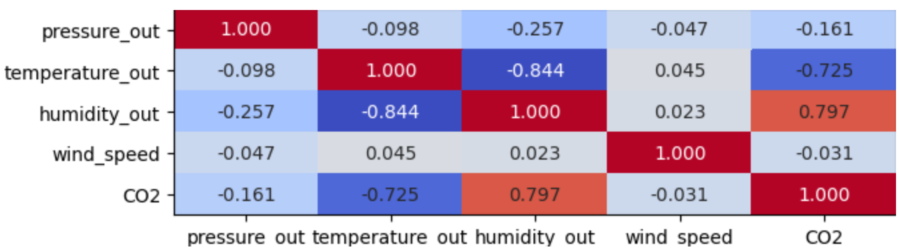


Figure 5.10: Correlation matrix of the outdoor parameters in August

In October, indoor temperature was found to have the highest correlation coefficient of 0.25. This correlation value is relatively lower compared to the value observed in August (Figure 5.11). Only those parameters that exhibit significant correlations were included in the development of the ANN models for each month. While computationally inexpensive, the decision to include only parameters with significant correlations was based on the idea that non-informative attributes, as assessed through correlation, have the potential to decrease the model's effectiveness and introduce uncertainty. This is especially crucial to consider when dealing with a limited number of data points, particularly when dealing with numerous attributes. It is a factor that can significantly impact model performance. Moreover, it is worth noting here that the correlation-based approach was used for parameter selection in the ANN models because feature importance analysis is only applicable to the RF models.

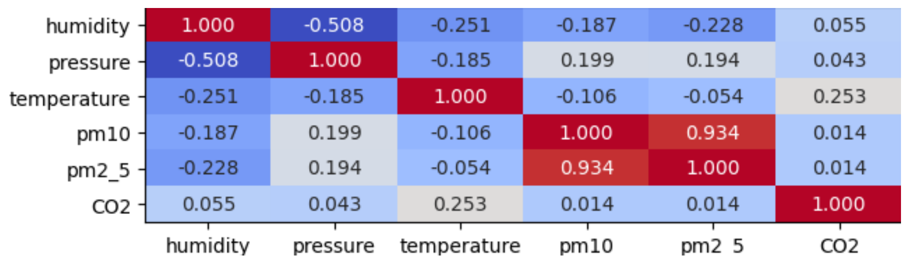


Figure 5.11: Correlation matrix of the indoor parameters in October

For the multivariate ANN models, two line graphs were generated to compare the predicted CO_2 levels against the actual CO_2 levels for the months of August and October. The x-axis of each graph represents the sample hours and the y-axis represents the CO_2 values. The RMSE for the August model was 16.574, which indicates that there is a relatively good fit between the predicted and actual values. Figure 5.12 shows a general trend where the predicted CO_2 levels followed the actual levels, with some fluctuations that could be attributed to random variations in the data.

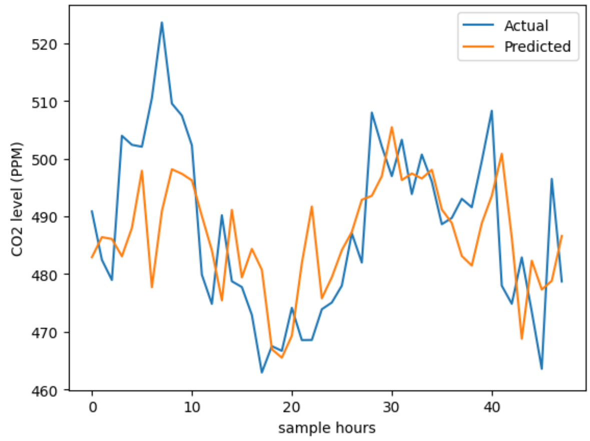


Figure 5.12: Sample hour prediction using the ANN model during August

For the October model, the RMSE was 15.642, which indicates a better fit when compared to the August model. The graph shows a clear trend where the predicted CO_2 levels followed the actual levels closely (Figure 5.13). There were some fluctuations, but they were less significant when compared to those in the August model.

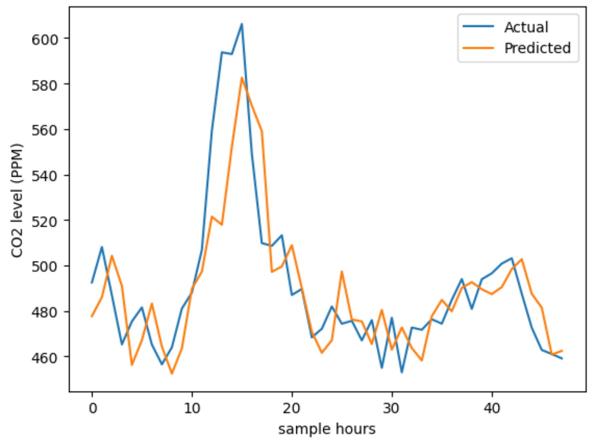


Figure 5.13: Sample hour prediction using the ANN model during October

The multivariate ANN models showed good performance in predicting the CO_2

levels, with the October model performing better than the August model. The correlation analysis that was used to select the model parameters seemed to have been effective in producing accurate predictions. However, it is important to note that these models were trained and tested using a relatively limited data set, and further testing and validation are necessary to assess their performance under different conditions.

In summary, this study has developed one-step univariate and multivariate RF and ANN models to predict indoor CO_2 levels in the Forum. The models were developed using data from two different months, August and October, and a number of parameters were used to determine their impact on CO_2 levels. The results showed that the most influential parameter for the univariate RF models was the CO_2 level from the previous hour, while the multivariate RF models highlighted the importance of indoor conditions (e.g., temperature and humidity). Meanwhile, the ANN models were based on highly correlated parameters: the August model incorporated indoor and outdoor temperature and humidity, while the October model used the CO_2 level from the previous hour and indoor temperature. The developed models serve as a baseline for the 24-hour ahead prediction of CO_2 levels, and can assist in identifying the parameters and sensing devices needed to understand the dynamics of indoor environmental conditions. By developing and comparing various models, this study provides insight into the predictive capabilities of different ML techniques and highlights the importance of selecting appropriate parameters for accurate predictions.

5.1.4 Multistep Prediction Models (24 Hour-ahead)

This section will describe the development of multi-step ML-based prediction models for CO_2 levels. There are two main approaches to developing a multi-step time series prediction model: recursive and direct forecasting. Recursive prediction uses a single model to predict one future value at a time. It then uses the predicted value to make the prediction for the subsequent time step. This process is repeated for each future time step that needs to be predicted. Although this approach is easy to implement and is capable of handling highly variable data, it can accumulate errors over time. This leads to a less accurate model, especially when the number of steps increases. Meanwhile, the direct approach, in which a separate model is used to predict the value for every hour in the day (24 models in total). However, this approach can be more complex to implement, and it limits the amount of data that are available to train and validate each model. Generally, the choice between recursive and direct prediction depends of the characteristics of the data and the requirement of the problem being considered.

The analysis of hourly CO_2 levels for the month of October is visualised through two sets of figures. The first set consists of 24 figures, one for each hour of the day, and displays the CO_2 profile over the entire month (Figure 5.14). The x-axis denotes calendar days starting from the 1 October and continuing to the end of the month, while the y-axis represents the CO_2 levels. Each figure allows for an examination of CO_2 levels in a specific hour of the day over the entire month. Upon careful examination, no clear pattern or trend was observed in the data for any hour of the day. This absence of a discernible pattern could potentially pose a challenge when developing a multi-step time series prediction model. It also underscores the need for techniques that can handle high variability and changing patterns, such as recursive predictions. To gain further insights into the distribution of the CO_2 levels in each hour, the second set of figures consisted of 24 histogram charts, one for each hour (Figure 5.15). Each figure shows the distribution of CO_2 levels in each hour over the month of October. Analysis of these charts reveals that there was no clear distribution for CO_2 levels in any hour of the day. This observation underscores the need for an approach that does not make a specific distribution assumption. Drawing on the findings of the previous analysis, it is apparent that the recursive approach represents the optimal strategy for predicting future CO_2 levels.

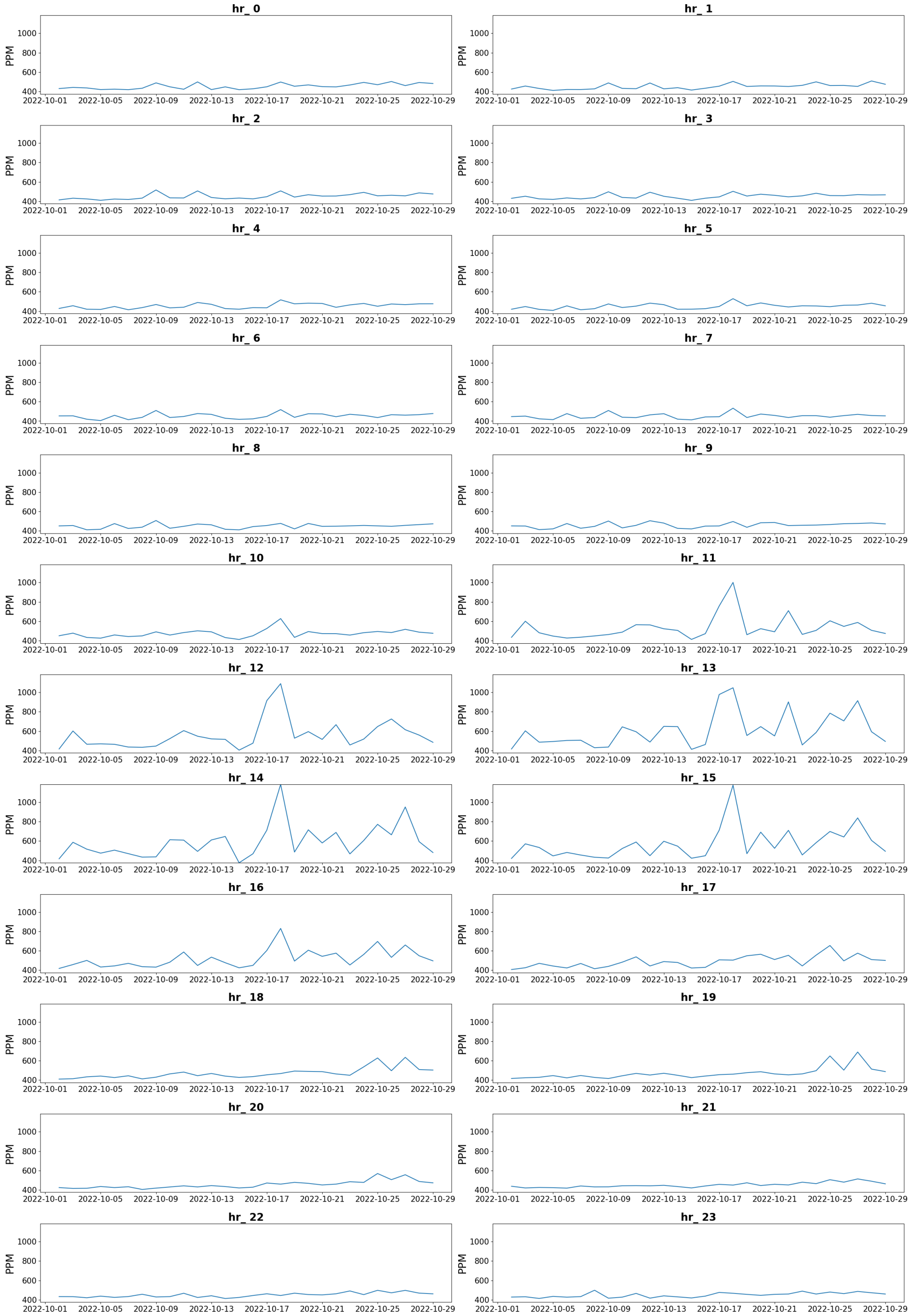


Figure 5.14: CO_2 profile for each hour of the day during October

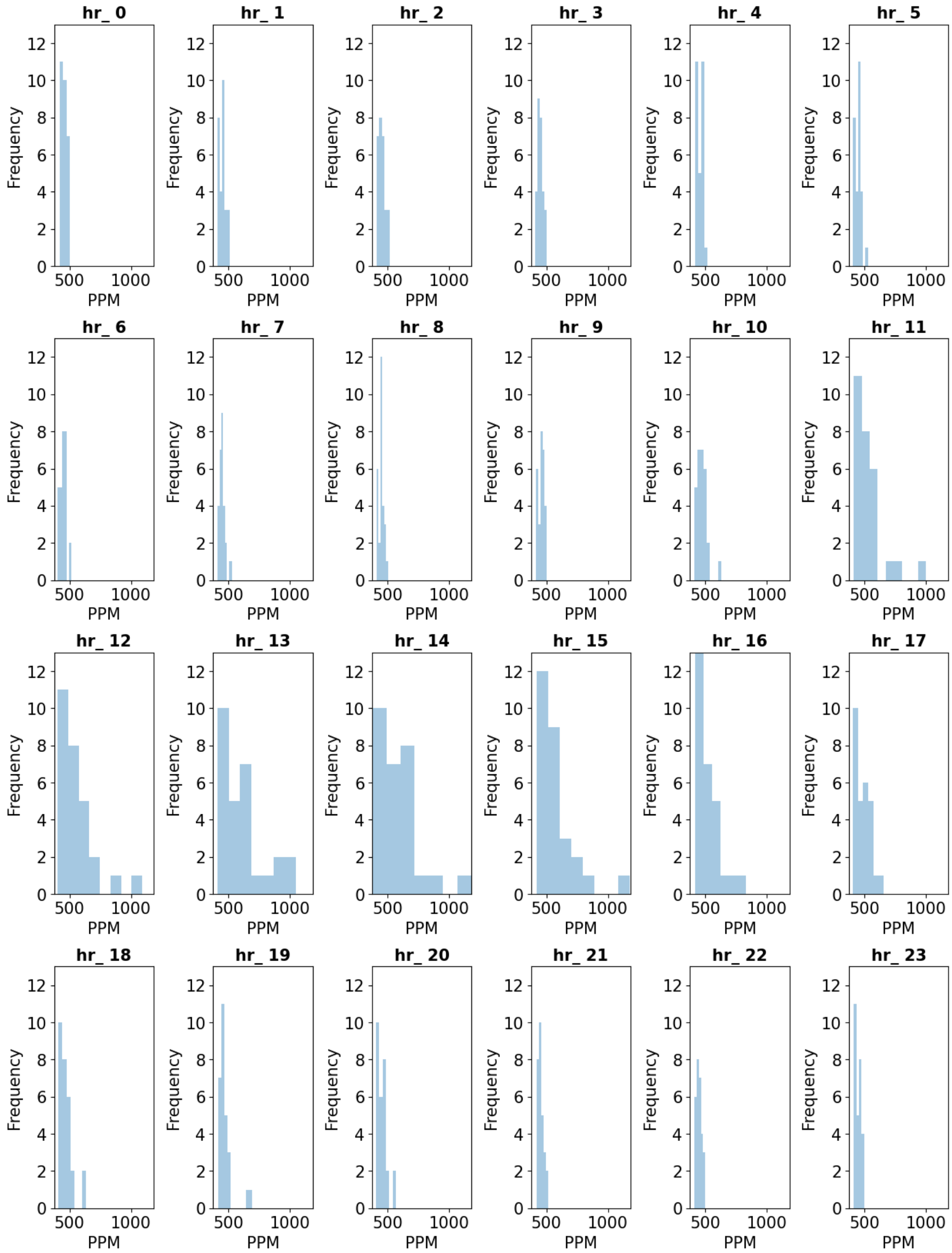


Figure 5.15: Distribution of CO_2 observations during October for each hour of the day.

Figure 5.16 displays the RMSE for the best performing recursive ML model (RF model) used for 24-hour ahead prediction of CO_2 levels. The x-axis represents the hours of the day (24 hours), while the y-axis represents the corresponding RMSE values. A clear pattern is evident in this figure. The RMSE values are generally low for the early hours of the day (1 am to 9 am), with values well below 30. As the day progresses, the RMSE values increase. The highest values are observed in the afternoon hours, where values exceed 250. However, the model's prediction error sharply decreases by the end of the day, after 7 pm. The observed RMSE pattern and the low RMSE values for the early and late hours of the day may be attributed to a more stable and predictable CO_2 pattern during these hours. In contrast, the observed increase in RMSE values during the afternoon may be a result of the higher variability and unpredictability of the CO_2 levels during this time, which lead to predictions that are less accurate. In addition, this pattern suggests that the model's performance may be influenced by the time of day and the associated characteristics of the CO_2 levels during that time.

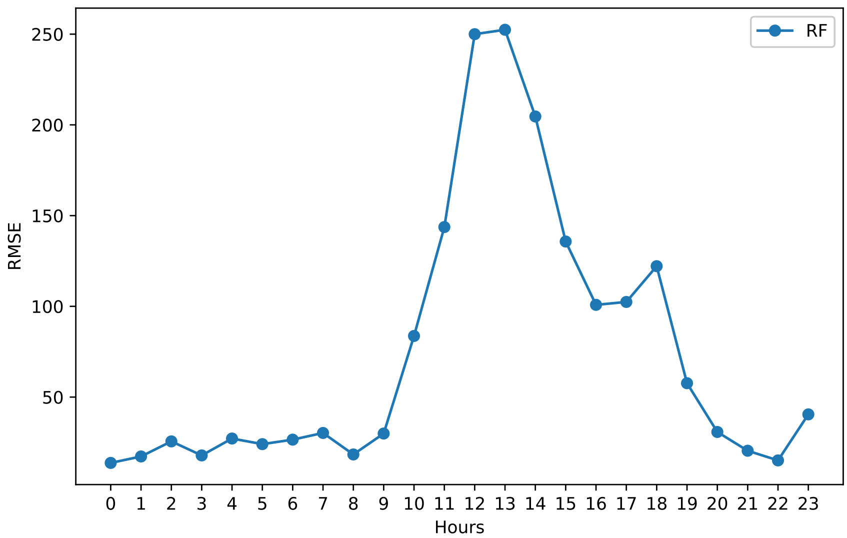


Figure 5.16: 24 hour-ahead CO_2 prediction performance of the RF model

Furthermore, given that the model predicts the next 24 hours, and that only one month of CO_2 level data is available for training and validation, the amount of data that is available for the model is relatively limited. This could potentially impact the model's accuracy, especially during hours of high variability.

5.1.5 Energy Model: Baseline

Energy modelling is an essential tool for designing and evaluating building performance, and Autodesk Revit is one of the leading BIM software tools that is used for this purpose. While Revit's add-ins provide basic energy analysis capabilities, third-party software tools such as EnergyPlus are often preferred for their comprehensive energy models, admissibility, and building code compliance checking. A number of options are available to prepare a BIM model for EnergyPlus simulation, including exporting Revit models as Industry Foundation Classes (IFC) or Green Building Extensible Markup Language (gbXML) files through add-ins or python conversion libraries. However, these options often suffer from compatibility issues, duplication of building elements and services, and missing information during the exchange process. For this reason, the Forum building's original model was exported from Revit as a gbXML file with only basic geometry information and was then further developed using the DesignBuilder tool. Additional information (e.g., construction material, occupancy schedule, HVAC system, and operational schedules) was incorporated to create a reliable energy model. It is important to note that for this study, only the Forum was considered, which was modelled as a single uniform zone, based on several interviews with the facility manager, a walk-around audit, and available mechanical system documentation.

The Forum uses a combination of a mechanical ventilation system consisting of an air handling unit (AHU) without a cooling function and natural ventilation using operable windows. It was assumed that the occupancy schedule during weekdays remains constant and reaches maximum space capacity in the afternoon, without giving consideration to the changing occupancy patterns. The occupancy schedule was assumed to be from 8 am to 5 pm on weekdays. Consequently, the mechanical ventilation system was modelled to operate from 8 am to 5 pm, aligned with the occupancy schedule. Moreover, the AHU contains a heater battery to regulate the temperature of the outside air before it is mixed with the indoor air.

The resulting baseline energy model must be calibrated to meet ASHRAE's requirement for calibrated energy models. However, measured energy consumption data were not available during this study. Therefore, the CO_2 profile from the simulation model was compared to the CO_2 from sensors to validate the baseline model. The simulated CO_2 and measured CO_2 profiles were compared using a line graph (see Figure 5.17). The x-axis represents calendar days during October and the y-axis denotes the CO_2 levels. The two profiles exhibited similar patterns, with both increasing, peaking, and decreasing at the same time.

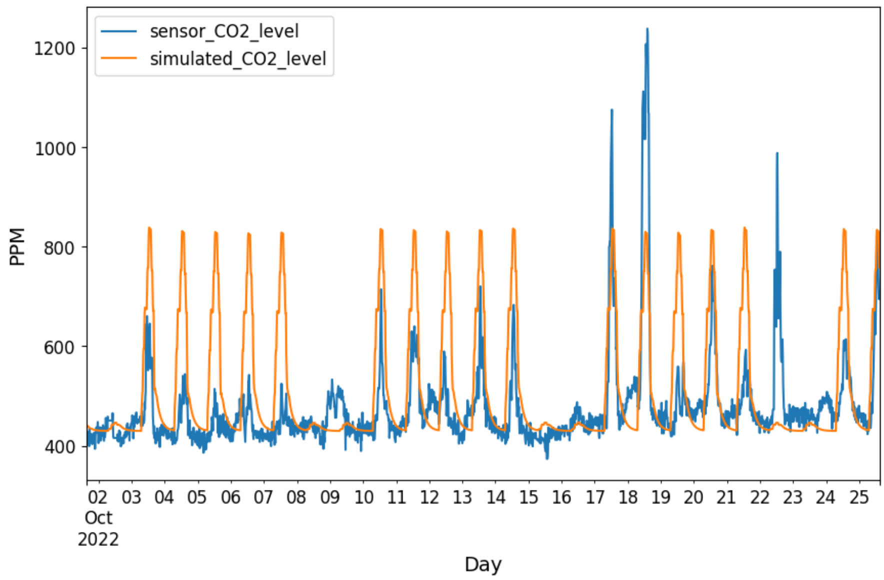


Figure 5.17: CO_2-based model calibration

Nonetheless, it was evident that the simulated CO_2 profile deviated from the measured CO_2 during the first two weeks of October, while the measured CO_2

went higher than the simulated values in the second half of the month. The observed discrepancies could be attributed to the modelling assumption that all weekdays have the same occupancy level and the fact that the first two weeks were at the start of the autumn semester, when the students had just returned to the university.

5.2 Energy Optimisation

The mechanical ventilation systems that are used to maintain indoor air quality consume significant amounts of energy, which has both environmental and economic implications. Therefore, efficient operation of the mechanical ventilation system can be an effective way to minimise energy use, while ensuring a healthy indoor environment. Therefore, this study proposed an optimisation strategy for the mechanical ventilation system, using GA, energy simulation, and ML techniques.

5.2.1 Optimisation Strategy

The first step in developing the optimisation strategy was to identify the problem parameters and formulate the objective functions. The ventilation system, which has an AHU, was analysed to determine the design parameters and identify the constraints. The AHU is formed of two axial fans with a single speed and a heater battery to regulate the temperature of the outside air. The energy consumption is mainly attributed to these two components. Although energy can be saved using an on/off strategy, this was not considered to be a good option because it can cause wear and tear, and reduce the lifespan of the fans. Additionally, frequent on/off switching could result in an increased energy consumption due to the power surge to start the system. Therefore, in the scenario presented in this study, once the fans are turned on, they stay on until they are scheduled to be turned off.

Controlling the supplied outside fresh air was considered to save energy, which can be adjusted through the recirculation rate, because the returned air can be put back into the system, thus reducing the required outside air volume and using less heating energy. Therefore, the percentage of the outside air to the total supplied air was considered as the design variable in the optimisation process. Therefore, this study aimed to minimise energy consumption while ensuring that indoor CO_2 levels remain within acceptable limits by adjusting the recirculation rate.

Two potential options were available to evaluate the objective functions. One was to start the optimisation process and then run the fitness evaluation by running EnergyPlus. However, this option was not practical considering the number of solutions in each population of the GA and the number of iterations that EnergyPlus would need to go through before generating results. In the second option, which was employed in this study, multiple scenarios were generated within EnergyPlus, each representing a different recirculation rate, with associated energy consumption and CO_2 concentration (Figure 5.18). A ML-based model was trained on the data resulting from the simulation scenarios and was used to predict two outputs (i.e., energy consumption and CO_2 level). The resulting ML model was used by the GA as a fitness evaluation to obtain the optimal solution. In this regard, the GA generates a population of potential solutions, each represented by a set of design variables. In this study, there is only one GA design variable, which is the fresh air percentage. The ML model is used to evaluate the fitness of each solution. The outputs of the ML model are returned to the GA, where a sequence of steps are carried out (e.g., selection, crossover, and mutation operators) to obtain an optimal solution over many generations. Using an ML model as a fitness evaluation function provided a more efficient and practical approach to optimisation when compared to directly using EnergyPlus for fitness evaluation. This happens because once the ML model is trained and validated, it can be used to predict the performance of the ventilation system accurately and in real-time without the need for time-consuming EnergyPlus simulations.

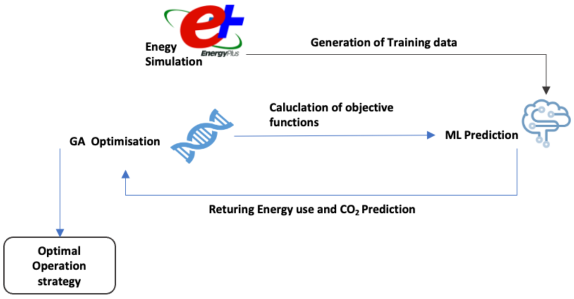


Figure 5.18: Workflow of the optimisation procedure integrating GA, EnergyPlus, and ML model.

5.2.2 ML Model

This section aims to develop an ANN model to predict energy consumption and indoor CO_2 levels based on energy simulation outputs. ANN was chosen over RF because of its ability to predict two different outputs with a single model, while RF requires two separate models. Additionally, the ANN model that is used in this section includes a different set of variables than the previous models because the problem can be seen as an approximation problem rather than a forecasting problem, which was the aim of the previous models. The input variables for this step were the state of the mechanical system (a binary variable where 1 represents ON and 0 represents OFF), outdoor temperature, fresh air percentage, and hour of the day. The reason for incorporating the 'state' variable is to allow for adaptability to changes in the operational schedule. While it is true that 'on' corresponds to energy use and 'off' signifies no energy use, the introduction of the 'state' variable permits the optimisation model to account for potential variations or shifts in the operational schedule.

In this study, the primary aim was to optimise the mechanical ventilation system by controlling the recirculation rate. This variable influences the CO_2 levels within the indoor space. Unfortunately, implementing direct control over recirculation rates was not feasible within the scope of the case study. Given this limitation, simulation emerged as the most viable approach to capture the dynamic impact of recirculation rate adjustments on CO_2 levels. By leveraging simulation, it was possible to model the interplay between ventilation strategies and CO_2 levels. Furthermore, the actual sensor data were obtained under typical operating conditions, which did not include the specific adjustments to recirculation rates that were central to the current study. Hence, relying solely on actual sensor data for model training would not have captured the full spectrum of scenarios relevant to the objective of the optimisation model.

The best ANN architecture was obtained by controlling several model parameters, such as the number of neurons, the optimiser selection, the transfer function, and the number of hidden layers. The latter parameter was considered in this model due to its complexity because it was configured to predict two target variables simultaneously. The final ANN architecture had two hidden layers with nine and two neurons for the first and second layer, respectively (Figure 5.19). All of the input variables were included in the final model.

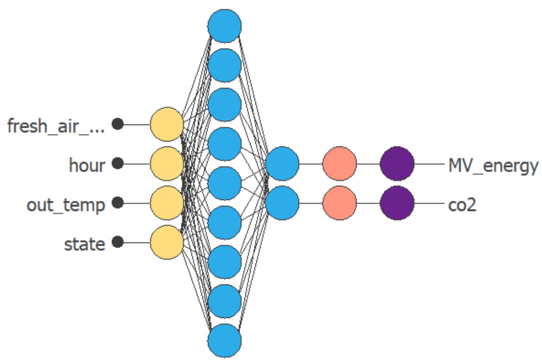


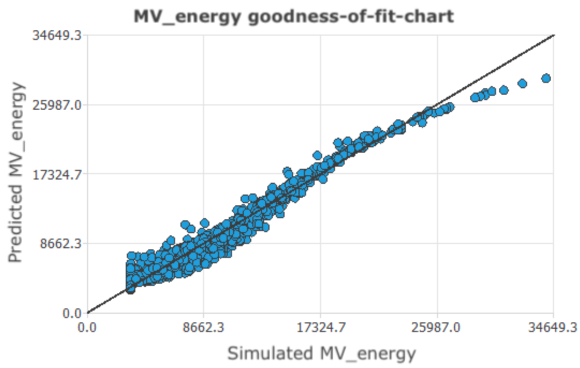
Figure 5.19: ANN model architecture for the prediction of energy consumption and CO_2.

The performance of the final ANN model was evaluated using the coefficient of determination (R^2) for both energy consumption and CO_2 levels. The R^2 for energy consumption was 0.9894, while the R^2 for CO_2 levels was 0.6509. Figure 5.20 shows that there is a very strong correlation between the predicted and simulated values for energy consumption, while a weaker correlation was observed for CO_2 levels. The discrepancy between these two predictions can be attributed to three key factors. First, the chosen input variables might have influenced the model's performance. This was evident when looking at the correlation coefficients between inputs and outputs (Table 5.3). The state input variable had the highest correlation value with energy consumption of 1, while the correlation value was 0.565 with CO_2 levels. The other input variables had roughly similar low correlation values with both outputs.

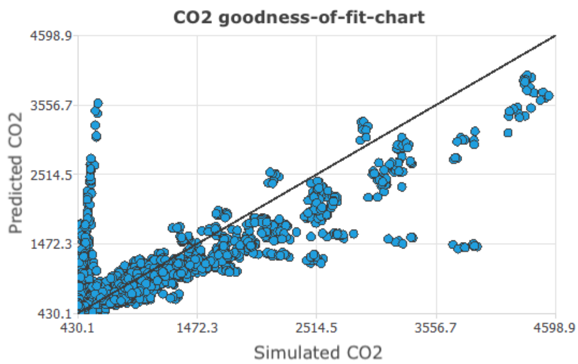
Second, potential bias in the training data towards higher CO_2 levels may lead the model to predict higher values even in scenarios where simulated levels are lower. In the provided goodness-of-fit chart (Figure 5.20-b), a comparison is made between simulated CO_2 levels and those predicted by the ANN, yielding an R^2 value of 0.6509. A notable phenomenon emerges at the initial segments of the chart, where the ANN tends to overestimate the CO_2 levels, especially at lower values. Given that the ANN was trained on simulated data, it would be expected that the predicted CO_2 levels closely align with the simulated ones. In essence, if the training data does not fully represent the variability in the simulated data, particularly at the lower CO_2 levels, the model might struggle to make accurate prediction in those regions. Third, the complexity of the model introduces a trade-off. The fact that the same model predicts both CO_2 levels and energy consumption might contribute to the observed discrepancy. There is a possibility that the model prioritises predicting energy consumption over CO_2 levels. This hypothesis is supported by the notable high R^2 value of energy consumption prediction, standing at 0.9894. Addressing this issue could involve re-evaluating the model architecture, fine-tunning feature selection, and augmenting the training data to include more examples of lower CO_2 levels.

Table 5.3: Energy consumption and CO_2 correlations with input variables

Input	**Energy consumption**	CO_2
State	1	0.565
Hour	0.101	0.308
Outdoor Temperature	-0.0768	0.171
Fresh air %	0.239	-0.389



(a) Energy consumption



(b) CO_2

Figure 5.20: Goodness-of-fit plot for the ANN model predictions of energy consumption and CO_2.

To compensate for the weak performance of the current model in predicting CO_2 levels, the limit of CO_2 levels was set to a lower value than the allowable limits by many standards, including ASHRAE and CIBSE. This will be further discussed when discussing the GA implementation.

It should be noted that the predicted CO_2 value in this model is different from the prediction in the previous section. In this model, the goal is to obtain the associated CO_2 level when changing the recirculation rate. Given that both energy and CO_2 data are simulation-based, the only way to evaluate the impact on CO_2 is by predicting it in the same way that the energy was obtained (i.e., through a model that was developed using simulation data). Currently, based on the proposed approach, the real impact of changing the operation settings cannot be captured with real data. Further clarification on this point will be provided when discussing the GA implementation.

5.2.3 GA Implementation

The GA was used to obtain the optimal operation strategy for the next 24 hours. The GA parameters were optimised to obtain the best results, including population size, selection, mutation, and crossover functions. These parameters are shown in Table 5.4. The decision variable, which is the recirculation rate, is allowed to vary from 0 to 100. The recirculation rate, denoting the proportion of indoor air that is recirculated through the system, is allowed to vary between the maximum value of 100, where no outdoor air is introduced into the system, and the minimum value of 0, where the indoor return air is completely exhausted from the system.

Table 5.4: Genetic Algorithm settings

Parameter	**Value**
Population size	100
Sampling method	Integer random sampling
Selection method	Tournament
Crossover method	Simulated binary crossover
Mutation method	Polynomial
Number of generation	1000

While the primary aim is to minimise energy consumption, the optimal approach may lead to high CO_2 levels, and therefore a constraint for CO_2 levels was incorporated into the optimisation algorithm. Although there is no clear and definitive limit for indoor CO_2 concentrations in university spaces, there are guidelines from organisations such as ASHRAE and CIBSE that can be used to determine the CO_2 limit. For example, ASHRAE Standards 62.1 assumes a maximum CO_2 concentration of 700 ppm above outdoor levels [297], which translates to a total concentration of 1100 ppm, assuming a typical outdoor CO_2 concentration of 400 ppm. A concentration of 1200 ppm or lower is generally considered to indicate acceptable indoor air quality according to CIBSE Guide A [319].

Considering the limitations of predicting indoor CO_2 levels accurately, the constraint on the CO_2 level was considered to ensure the optimisation strategy does not result in excessively high indoor CO_2 levels. Specifically, an upper bound of 900 ppm was set, which is below the acceptable limits. The decision to use such a low threshold was twofold. First, the accuracy of the ANN model in predicting CO_2 levels was found to be limited. Thus, by setting an upper limit of 900 on the indoor CO_2 concentration, any solution that results in a CO_2 level above this threshold will be rejected by the algorithm. Second, the GA relies on the 24-hour ahead CO_2 predictions that were generated by the models in Section 5.1.4 to determine the optimal operation strategy. However, the best-performing model resulted in significant prediction errors. In light of these considerations, a CO_2 limit of 900 ppm was selected to ensure that indoor air quality is maintained while minimising energy consumption.

Fitness Evaluation

The fitness evaluation process follows the workflow outlined in Figure 4.4. To optimise the operation of the system for a 24-hour period, the GA algorithm generates an initial population with varying recirculation rates. The optimisation loop is then initiated, with outdoor temperature, CO_2 level (as predicted in section 5.1.4), and state provided as inputs for each hour. This results in an input matrix with 24 rows and 3 columns. The ANN model uses this input to predict energy consumption and the associated CO_2 level. Solutions that result in high CO_2 predictions are rejected, and the best solutions in the current generation continue through the GA cycle. This process is iterated until the termination criteria are met, which was set at 1000 generations, at which point the optimal operation strategy is generated.

Results

The optimisation strategy that is implemented in this study resulted in a significant energy saving of approximately 35% compared to the baseline scenario generated by EnergyPlus. A set of three subfigures was created to present the energy consumption, outdoor temperature, and optimal recirculation rate over the course of the month of October (Figure 5.21). The x-axis of all of the subfigures represents the 24-hour interval.

The first subfigure gives the energy consumption profiles for both the optimised and baseline scenarios. The baseline scenario implements a typical ventilation strategy, which is maintaining a ventilation rate of 3 $L.s^-1$ per person during the occupied period (8 a.m. to 5 p.m.).

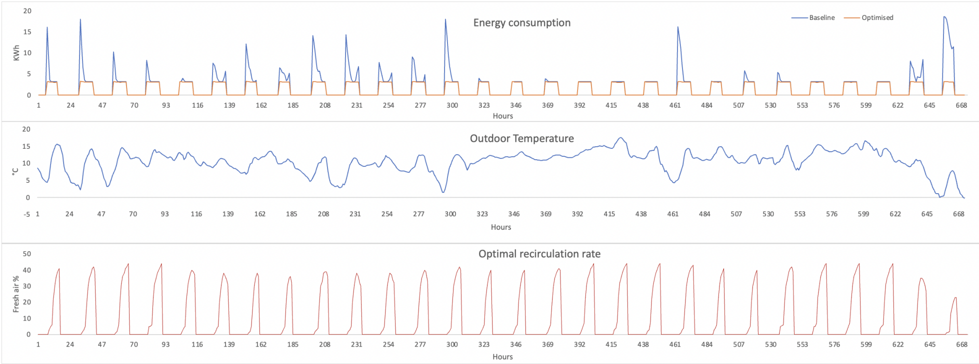


Figure 5.21: Optimisation strategy results showing energy consumption, outdoor temperature, and optimal recirculation rate.

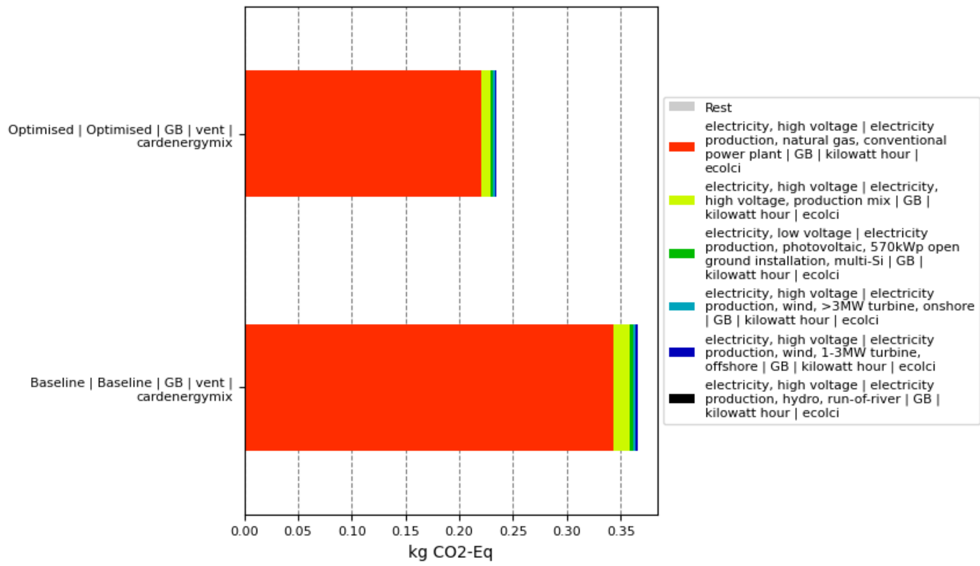
It can be seen that there is a spike in the baseline energy consumption demonstrates the early hours of each day, which can be attributed to the low outdoor temperature during the morning hours. Without the ability to control the percentage of fresh air supplied to the space, additional heating energy is required. In contrast, the optimised scenario shows a slight increase in energy consumption during the early hours, which is significantly lower than the baseline scenario. This happens because the optimised scenario controls the supply air based on the indoor CO_2 concentration during the early hours. As Section 5.1.1 has shown, the indoor CO_2 concentration is low during this time and is near the outdoor CO_2 levels, and hence no additional heating energy is consumed. This is the reason why the optimised scenario outperforms the baseline scenario during this period. However, both scenarios demonstrate similar energy consumption as the outdoor temperature increases throughout the rest of the day. The second subfigure shows that the outdoor temperature is consistent with the energy consumption profiles, with the temperature being lower during the early hours of the day and increasing as the day progresses. Finally, the third subfigure shows the optimal recirculation rate throughout the month of October. The fresh air supplied to the indoor space demonstrates a similar pattern to the indoor CO_2 levels (i.e., increasing throughout the day as the CO_2 concentration levels increase due to space occupancy). The resulting optimal recirculation rate validates the optimisation results and shows that the optimised scenario is valid.

5.3 LCA Results

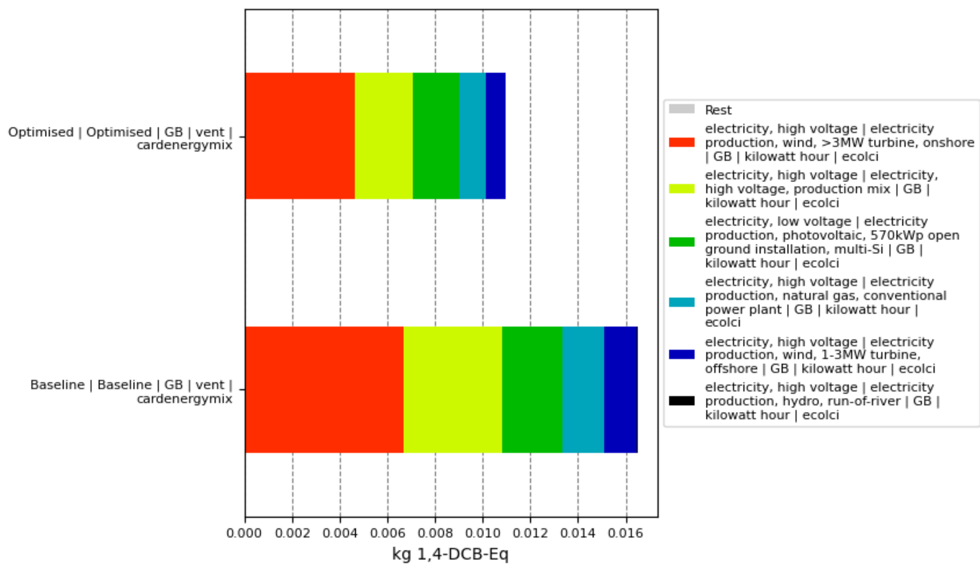
This section discusses the LCA results of the two scenarios (a baseline scenario and an optimised scenario) based on four impact categories: climate change, human toxicity, fossil fuel depletion, and metal depletion. The LCA results are presented in Figure 5.22, which illustrates the contribution of the unit processes to each impact category for each scenario.

This figure shows that the baseline scenario has a higher impact than the optimised scenario across all of the impact categories. In particular, for the climate change impact category, the baseline scenario has a significantly higher impact than the optimised scenario. The primary contributor to this impact category is electricity production from natural gas, which is the primary source of electricity in Wales and accounted for approximately 63% of the total. However, the impact of climate change from other unit processes is relatively small, due to their smaller share of overall electricity production.

These findings highlight the importance of optimising energy performance in buildings in the short term because it can significantly reduce the impact of climate change. For the fossil fuel depletion impact category, the results also show that the optimised scenario outperforms the baseline scenario. However, when considering the metal depletion impact category, the unit process that contributes the most is onshore electricity production, mainly due to the large amount of raw materials that are required to build wind turbines. For the human toxicity impact category, the highest contributors are electricity production from onshore wind and solar. Optimising energy consumption directly reduces the demand for

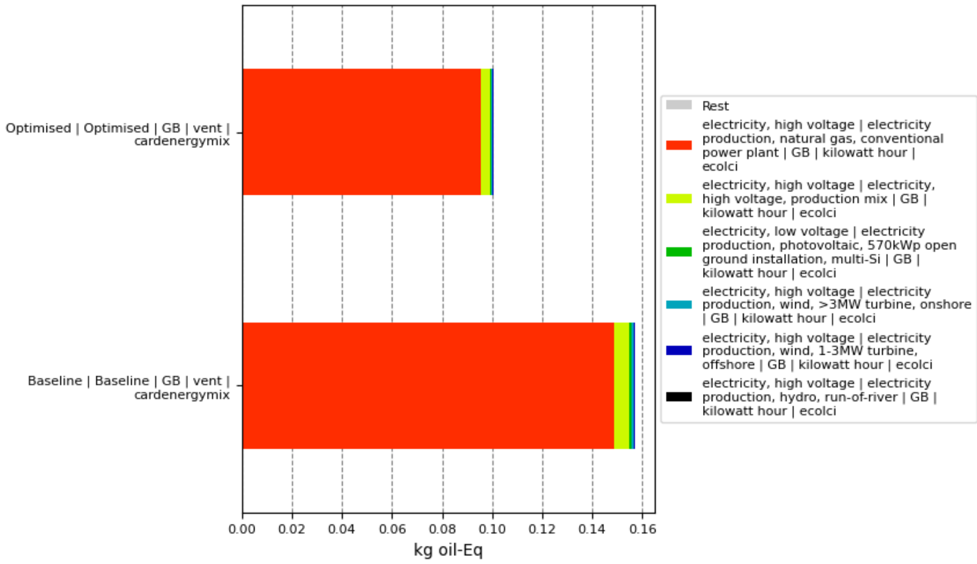


(a) Climate change

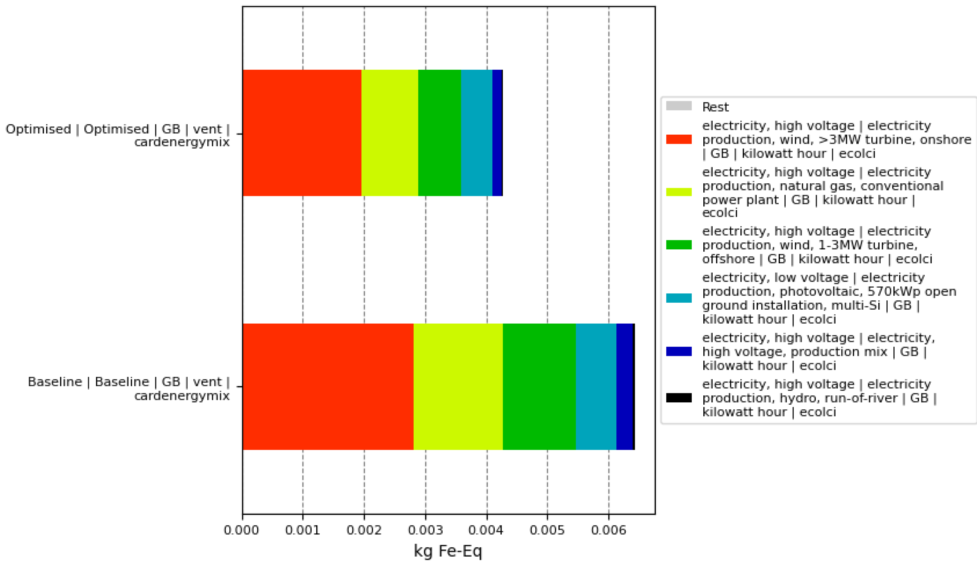


(b) Human toxicity

Figure 5.22: Comparison of baseline and optimised scenarios under sample impact categories (climate change, human toxicity, fossil fuel depletion, and metal depletion.



(c) Fossil fuel depletion



(d) Metal depletion

Figure 5.22: Comparison of baseline and optimised scenarios under sample impact categories (climate change, human toxicity, fossil fuel depletion, and metal depletion.

energy from the grid. This, in turns, lowers overall resource extraction, including metals used in energy production, and reduces the production of hazardous and toxic materials associated with some energy technologies, such as PV panels.

These results highlight the importance of enhancing the energy performance of buildings. Even when the transition towards cleaner and renewable energy technologies is made, their environmental impacts will remain a challenge. Therefore, reducing the energy consumption of buildings should be considered as a viable strategy to mitigate their environmental impact.

5.4 Discussion

This section delves into further environmental considerations regarding the proposed framework. While the framework exhibits considerable potential in enhancing energy efficiency and minimising the environmental impact of buildings, it is essential to recognise the environmental footprint associated with its resources. This discussion encompasses the concept of GREEN IT/S, the computational costs of ML training, and the environmental implications linked to connected objects. Additionally, the discussion extends to address the vital aspects of scaling up the framework, outlining the steps necessary for its adaptation across a broader context. Furthermore, the discussion includes the integration of natural ventilation as a sustainable design option.

5.4.1 Further Environmental Considerations for the Proposed Framework

The developed framework represents an important step towards promoting sustainability within the built environment. By leveraging the progress in computation and data-capturing technologies (e.g., sensors), the framework has the potential to enhance energy efficiency, thereby reducing the environmental impact of buildings during the operational phase. However, it is crucial to recognise that these resources bear their own environmental footprint, stemming from energy consumption for operation and maintenance, utilisation of natural resources for electronic hardware manufacturing, and the eventual disposal of electronic waste at the end of their useful life. Therefore, the fundamental concern lies in optimising these advancements in a sustainable and responsible manner.

Various sectors of society, ranging from government bodies and corporations to individuals and scientific community, have become increasingly conscious of the environmental dimensions of information technologies and systems [320], encapsulated by the term GREEN IT/S. GREEN IT predominantly pertains to the environmentally conscious design and management of hardware and IT infrastructure, while GREEN S embodies the enhancement of information flow and management [321]. Numerous initiatives have been established to advance the practice of GREEN IT/S, with a primary focus on enhancing the energy efficiency of IT infrastructure, reducing electronic waste, and advocating for equipment recycling and reuse. One noteworthy example is The European Green Digital Coalition (EGDC)[1], which stands as a consortium of companies dedicated to advancing both the green and digital transformation of the EU. The mission of the ECDG is to maximise the sustainability benefits derived from digitalisation, thereby supporting the EU in achieving its climate and digital objectives. The coalition places emphasis on investing in green digital solutions that conserve energy and materials, collaborating with non-governmental organizations and experts to develop metrics for assessing the environmental and climate impact of green digital technologies, and working across sectors to formulate recommendations and guidelines for achieving green digital transformation, hence, benefiting the environment, society, and the economy. Another prominent initiative is the Climate Savers Computing Initiative, which champions practices that minimise

[1]https://digital-strategy.ec.europa.eu/en/policies/european-green-digital-coalition

the environmental footprint of computers and electronic devices[2]. This includes efforts to reduce energy consumption, promote the use of recycled materials, and advocate for responsible disposal of electronics. These initiatives foster a collective commitment towards integrating sustainable practices into IT operations.

Topics within GREEN IT/S domain include energy efficiency in cloud computing [322]. The adoption of cloud computing holds the potential to significantly reduce global data centre energy consumption, enhance power usage efficiency, promote recycling efforts, and minimise water consumption for cooling purposes [323]. Sustainable procurement is another critical aspect, emphasising the acquisition of IT products and services that align with environmental and social responsibility [324]. This entails a holistic evaluation of the environmental impact of a product or service throughout its entire life cycle. Additionally, electronic waste reduction is of paramount importance, entailing strategies such as prolonging the lifespan of IT equipment and implementing recycling and refurbishing protocols when equipment reaches the end of its operational utility [325]. In light of the proposed framework and the scope of this study, the discussion will now focus on two crucial aspects that are integral to the implementation of the framework: the computational cost associated with ML training, primarily in terms of energy consumption, and the environmental considerations regarding connected objects (i.e., sensors).

Computational Cost

The computational cost associated with generating real-time strategies is an important aspect, which has not been factored into the current study. In particular, the retraining frequency of ML models and the associated energy consumption. The frequency of retraining hinges mainly on two factors. First, the availability and characteristics of new data and whether the inclusion of new data can improve the model performance. Specifically, in the context of this study, retraining

[2] https://www.climatesaverscomputing.org/

is needed when new data is believed to have seasonal variations that were not adequately captured by the previous dataset. This phenomenon is commonly referred to as data drift [326]. In essence, it means that new data have characteristics that deviate from established patterns. Second, an important factor for determining when retraining is needed is if there is a noticeable drop in the model performance [327]. It is considered a good practice to monitor the performance of ML models and carry out retraining when deemed necessary, which can be achieved by tracking the performance over time. However, it must be noted that retraining can be both time-consuming and computationally expensive, hence, careful evaluation of the benefits against the associated costs is required before conducting model retraining.

ML researchers are increasingly considering energy consumption in ML computations, but most research still prioritises the performance of ML models without regard to computational cost [328]. The energy usage of ML models is influenced by several factors, including the choice of hardware used for ML training [329]. A study demonstrated that specialised processing units, such as GPUs require less energy compared to more general-purpose CPUs [330]. Furthermore, the selection of algorithms for training the model influences the energy requirements for ML computations [328]. A third factor pertains to the size of the dataset and the complexity of the model deployed [331]. A further consideration is needed with regard to data acquisition, specifically when determining whether to store data in centralised data centres or on local servers. This strategic decision has significant implications for the energy demand associated with maintaining IT infrastructure.

Environmental aspects of monitoring systems

While the environmental monitoring system integrated into the framework holds the potential to reduce the environmental impact of buildings, it is imperative to broaden the scope of consideration beyond the direct impact encompassing the hardware's life cycle (i.e., extraction of raw materials to end-of-life treatment). This expansion should not only consider the energy consumption of the sensors but also the energy demand of other essential components in the network, such as gateways and servers [332]. This holistic perspective ensures a thorough understanding of the environmental footprint associated with employed monitoring systems, ultimately providing a foundation for devising strategies to further mitigate the overall environmental impact of the IT infrastructure.

5.4.2 Scaling Up the Proposed Framework

This section delves into the critical considerations for expanding the scope of the proposed framework. Scaling up encompasses various dimensions, each playing an important role in ensuring the seamless application of the framework across a broader context. From data collection to optimisiation and ML models, and addressing building infrastructure, the key steps required to adapt the framework for a larger scale and longer time frame are explored. These considerations are imperative for maximising the applicability of the proposed framework in diverse building environments. The focus here is on navigating through the essential aspects of scaling up the framework to enhance its real-world effectiveness.

Data collection: Consideration of data collection and integration is imperative to effectively scale up the implementation of the proposed framework. In extending the scope, it is crucial to identify additional parameters that may not have been included in the current study. In this regard, two aspects must be considered. First, a thorough understating of the specific application or use case is paramount. For instance, parameters vital for optimising thermal comfort and HVAC systems may differ substantially from those pertinent to optimising lighting systems. To illustrate, temperature differentials, humidity levels, and airflow rates become critical for HVAC optimisation, whereas lux levels, and daylight levels are important parameters to consider for lighting optimisation. Second, consideration must be given to parameters that exhibit variability across different spaces and zones within the building such as room size, height, function, and layout. These considerations ensure that the solution can be applied seamlessly across diverse spaces to accommodate the unique requirements of each space.

Additionally, it's essential to carefully evaluate the granularity and frequency of data collection. The determination of the appropriate granularity hinges on the specific requirements of the use case. This entails thinking about questions of whether data should be gathered at minute intervals, 15-minute intervals, on an hourly basis, or at a different regular interval. This choice is contingent on the frequency at which building systems need to be actuated and the speed at which responses to changing conditions must be executed. Striking a balance between data granularity and how fast the system needs to respond ensures an optimal and resource-efficient approach to data collection and integration on a large scale.

Sensor placement and numbers: It is important to first identify the minimum number of sensors at the building level. This consideration is contingent on a comprehensive understanding of the operational context of the building and the specific parameters that must be captured. Moreover, the strategic placement and positioning of these sensors merit careful consideration. A thorough evaluation of the building's layout, and potential sources of variability is essential in order to position sensors in locations that provide the most representative data. Also, it is imperative to factor in the financial aspects associated with sensor acquisition and maintenance (i.e., the cost of acquiring sensors, along with the expenses linked to their maintenance). Furthermore, the energy consumption of the sensors must be weighed against the desired frequency of data collection. Striking a balance between the desired data collection frequency and the associated energy consumption is vital in order to maintain an efficient and sustainable smart infrastructure.

Scalability of ML models: Scaling ML models to the building level involves a careful examination of their adaptability and performance when faced with a more extensive dataset encompassing the entire building and a year's worth of data. In the context of the current study, this process entails assessing how well the models, which were initially designed for a single space and focused on two months, can be extended to cover a broader spatial and temporal scope. To achieve this, several key considerations come into play. First, the ML models developed in this study need to demonstrate a capacity for generalisation. They should be able to learn patterns and relationships from the localised data in a way that allows them to make accurate predictions when applied to a larger dataset across the entire building. Second, re-evaluating the features used in the models. It is crucial to verify that the selected features remain relevant and informative when applied to the broader context of the entire building. This involves considering whether additional features or parameters need to be incorporated to account for the diverse conditions and usage patterns across different spaces. Third, it is important to consider the computational resources required to train and re-train the models at a larger scale. This entails ensuring that the hardware and IT infrastructure can handle the increased data volume and the computational requirements of the ML models.

Optimisation algorithms: Utilising optimisation algorithms for a broader context involves assessing their effectiveness when applied to a larger-scale dataset and operational framework. In the context of the proposed framework, there are some key considerations for extending the scope of the optimisation model. First, it is important to evaluate whether the optimisation used in the localised implementation can be effectively extended to handle the entire building. This involves ensuring that the optimisation model is not overly specialised to the conditions of the Forum or the limited time frame. Additionally, it is crucial to revisit the parameters and settings used in the GA model to ensure that the chosen configurations can be applied to a broader spatial and temporal scope. Indeed, adjustments and fine-tuning of constraints may be necessary to adapt to the varying conditions throughout the building. Finally, similar to the scalability of ML models, considerations regarding computational efficiency and computational resources are warranted.

Building infrastructure: In the context of the proposed framework, it is crucial to address the building infrastructure to enable an effective implementation of the decision support system. For instance, actuators play a pivotal role in the physical implementation of optimisation strategies, converting control signals into physical actions. In this regard, upgrading the building's control system with actuators allows for a responsive adjustment to building systems such as HVAC components, and controlling window operations. Moreover, integrating the proposed framework with the existing Building Management System (BMS) is imperative to seamlessly implement the recommendations generated by the ML and optimisation models. Additionally, a robust communication network is essential for efficient information exchange and control signal transmission between various components of the building, including sensors, actuators, controllers, and the BMS.

5.4.3 Natural Ventilation

Although natural ventilation was not considered in the proposed framework, it is indeed an effective and sustainable building design option that has the environmental benefit of reducing energy consumption for indoor ventilation. The following considerations outline how natural ventilation could be integrated into the system and the potential challenges it may pose:

Climatic considerations: It is important to assess the prevailing climatic conditions. Natural ventilation is particularly advantageous in regions with moderate to warm climates, where the demand for space heating and cooling is limited. A thorough evaluation of the trade-offs between energy savings from natural ventilation and the energy demand for space conditioning is essential.

Building type and characteristics: The type and design of the building exert a substantial influence on the viability of natural ventilation. For instance, residential and office buildings, characterised by limited occupancy and access to windows, are prime candidates for natural ventilation strategy. Conversely, buildings with large open spaces, and diverse occupants, such as schools or facilities with strict indoor air quality requirements, such as hospitals, present challenges for effective implementation of natural ventilation. Additionally, unique building features, such as shape, orientation, opening configuration, and internal layout significantly affect natural ventilation effectiveness.

Control and operation of natural ventilation systems: The way in which natural ventilation systems are controlled and operated is a crucial factor to consider. While manually operated systems may present challenges for optimisation due to occupant behaviours and preferences, automated systems offer significant potential for maximising the benefits of natural ventilation.

In light of the above considerations, it is evident that a thorough evaluation is warranted before integrating natural ventilation into the proposed framework. Once these fundamental aspects are addressed, the framework can be extended to incorporate natural ventilation. To accurately simulate the impact of natural ventilation on indoor conditions, Computational Fluid Dynamics (CFD) simulation can be leveraged to model air change rates resulting from natural ventilation. This requires detailed information on space layout, occupancy profile, and window design features, including the size, type, and orientation. Multiple scenarios of window control can be simulated to identify influential factors that can be optimised, such as window state. Simulation for mixed mode ventilation is also required. Subsequently, the optimisation model utilised in this study can incorporate the outcomes from the CFD simulation to optimise the energy consumption of the mechanical ventilation system. It is imperative that the decision variables within the optimisation model include controllable parameters from both ventilation strategies (i.e., natural and mechanical) to ensure an effective and optimal strategy.

5.5 Summary

The primary focus of this chapter was to address the second and third research questions, which state:

Can access to dynamic data provide more accurate accounts of the environmental impacts during the operation stage?

How can machine learning and optimisation be leveraged to reduce the environmental impact of buildings?

To address the second research question, this chapter explored the potential of using dynamic data, namely sensor data, to enhance the understanding of a building's operation, and improve its energy and environmental performance. The main idea behind leveraging dynamic data was to contextualise it and then use it to improve the building's operation. This chapter has shown that leveraging sensor data enables the capture of the dynamic nature of buildings, leading to a more comprehensive understanding of their operational requirements, which in turn improves the accuracy of the LCA results.

The results show that CO_2 levels in indoor environments are highly dependent on factors such as occupancy, ventilation rates, and time of day. The data also highlight the need for a dynamic and flexible ventilation strategy that can be adjusted in response to changing conditions, rather than a fixed schedule that may result in over-ventilation or under-ventilation. By leveraging dynamic data, the building's managers can make informed decisions to adjust ventilation rates in response to changing conditions, which will lead to improved energy and environmental performance.

Several models were developed to predict the observed CO_2 levels in the case study site. Univariate models using only historical CO_2 levels and multivariate models incorporating additional variables (e.g., outdoor and indoor conditions, time of day, and day of the week) were developed using RF and ANN techniques. Feature importance and correlation analysis were used to identify the most influential parameters for each model. Furthermore, this study also explored the development of multi-step ML-based prediction models for CO_2 levels. The results of these models were discussed in detail throughout this chapter. Finally, the outputs of these models were incorporated in the subsequent step, which focused on the optimisation of the mechanical ventilation system.

To address the third research question, an optimisation strategy for a mechanical ventilation system was developed to reduce energy consumption, while ensuring healthy indoor air quality. This approach utilised a combination of GA, energy simulation, and ML techniques. Therefore, this study identified the design parameters and constraints of the ventilation system, with a focus on the AHU and heating system. The optimisation process aimed to adjust the percentage of outside air supplied to the system, which is a design variable that affects the recirculation rate, thereby reducing the required outside air volume and heating energy.

The optimisation process uses an ML model, namely ANN, which leveraged the energy simulation results to predict energy consumption and indoor CO_2 levels. The ANN model was trained on the simulation data and was then used by the GA as a fitness evaluation function to obtain the optimal solution. This approach provides a more efficient and practical method when compared to directly using energy simulation for fitness evaluation. The optimisation strategy resulted in a significant energy saving of about 35%, when compared to the baseline scenario.

The output of these models was used to inform the LCA model to evaluate the environmental impact of a building's energy consumption. By using accurate and reliable models, the LCA model can more effectively evaluate the environmental impacts of buildings, using factual data instead of relying on generic data that are based on typical building usage and operation. This contributes to the accuracy of the LCA model and provides valuable insights for decision-makers who wish to create more environmentally-friendly practices. This study also highlighted the potential of ML and optimisation techniques to reduce the environmental impacts of buildings by improving their energy efficiency.

Chapter 6

Semantically-enabled Life Cycle Assessment

This chapter aims to answer the research question *"Can a semantic web approach provide a sound basis to facilitate and streamline the applications of LCA in buildings?"* by showcasing how semantic modelling and interoperability can be utilised to automate and streamline the process of LCA in the building domain. The role of semantics in enabling data exchange between different components (e.g., BIM, sensors and LCA databases) will be explored to highlight its potential to overcome the challenges of conducting LCA studies in a complex and heterogeneous environment. It is pertinent to note that the semantic modelling presented here will be confined to the context of the proposed framework described in Chapter 4. This chapter is not intended to develop a semantic model for the entirety of LCA.

6.1 Background

LCA of the built environment is a complex and multifaceted task that requires the integration of multiple components and data sources. As demonstrated in the framework introduced in Chapter 4, LCA studies require the integration of various data models and entities (e.g., BIM, sensors and LCA databases), as well as modelling techniques such as simulation, prediction, and optimisation. The work conducted to gather relevant inventory data was manually carried out, which required a deep understanding of the activities involved in the LCA study. Furthermore, mapping the unit processes to the corresponding activities in the Ecoinvent database was a labour-intensive and time-consuming process. This challenge is further exacerbated by the fact that the resources that are used in LCA studies are structured in a certain way, with their own data structure. Moreover, the links and relationships between data points across different data sources are missing, and the relationships are implicit rather than explicit, which requires some level of expertise to make the connections. These issues were explored in-depth in the literature review, especially when the integration between building BIM, and LCA tools and databases was investigated. Moreover, LCA handles vast amounts of data, with multiple data providers and software vendors offering data in various formats; while some databases may be compatible with multiple LCA tools, others are designed for use with specific tools only.

To address these challenges, this chapter will showcase the importance of semantic modelling and interoperability in LCA, as well as their potential to streamline the application of building LCA. It will be argued that adapting LCA for a sustainable built environment will be limited and challenging without a semantically-enabled system. The use of semantics to orchestrate the interaction between different components of the LCA model can facilitate the exchange of information, reduce human intervention, and streamline the LCA process. Therefore, this study adopted an ontological approach to overcome the challenges of conducting LCA in the building domain by creating a shared conceptualisation and understanding of the involved fields, thereby enabling semantic modelling and interoperability.

The ontology community makes a clear distinction between two types of ontologies. One category primarily functions as taxonomies, while the other delves deeper into domain modelling, and imposes restrictions on domain semantics [333, 334]. These are referred to as lightweight and heavyweight ontologies, respectively. Lightweight ontologies include fundamental elements such as concepts, concept hierarchies, inter-concept relationships, and descriptive properties. Conversely, heavyweight ontologies extend lightweight ones by introducing formal axioms and constraints. In this study, the ontology developed for the framework pertains directly to its components and scope, with limited concepts and relationships, which aligns with the characteristics of a lightweight ontology. Furthermore, developing an ontology with complex axioms and constraints (i.e., heavyweight ontology) might be over-engineering for the purpose of the current study. Therefore, this study adopts a lightweight ontology to ensure that the ontology remains manageable, easier to understand, and adequate for depicting the essential concepts and relationships.

6.2 Ontology Development Methodologies

Developing an ontology for LCA-based building environmental performance requires a methodical approach that considers the unique requirements of the domain. A successful ontology must capture the diverse range of requirements that are involved in the LCA of an asset, such as enabling technologies, modelling tools, and LCA concepts. Uschold reported the most important design criteria for ontology development that were originally proposed by Gruber [283, 335], including clarity, extensibility, and reusability. The defined concepts must be clear and objective, and the definitions can be expressed informally using natural language or by means of formalism (i.e., logical axioms). In addition, to maximise subsequent reuse and extensibility, ontologies should be designed to strike a balance between specificity and generality. This requires the creation of an ontology that is specific enough to perform the intended task, while avoiding excessive specificity that would limit its usefulness to others.

Several methodologies have been proposed in the literature to aid ontology development, such as Grüninger and Fox's methodology [336], who were the pioneers of using questions as a means of evaluating ontologies, as proposed in their seminal work on the Ontological Framework for Enterprise Modelling in the TOVE project. Ontology-Driven Knowledge Management (ODKM) is another approach for developing and implementing ontologies for enterprises [337]. The purpose of ODKM is to aid organisations in creating knowledge-based management systems. This methodology distinguishes between knowledge meta-processes and knowledge processes, where the former facilitate ontology development and the latter facilitates ontology usage. Another approach is the UPON methodology, which applies software engineering principles to ontology development by incorporating widely-used standards, such as the Unified Modelling Language (UML) and the Unified Software Development Process (UP) [338]. This methodology is iterative and incremental, involving knowledge engineers and domain experts at each stage to achieve scalability and flexibility in the ontology's design. METHODOLOGY [219] has been recognised for some time in the realm of ontology development [339], and takes into account the entire ontology life cycle, from planning and conceptualisation to implementation and maintenance. METHODOLOGY emphasises that activities such as knowledge acquisition, documentation, and evaluation must be carried out in parallel throughout the entire ontology development process. Finally, the NeON methodology [220] is an extended version of METHODOLOGY that offers several advantages over its predecessor, which include simplicity, the provision of comprehensive documentation, and the use of a scenario-based approach to guide ontology development. Consequently, the ontology in this research has been developed using the NeOn methodology due to these advantages.

6.3 LCA-based Building Environmental Performance Ontology

This section presents a comprehensive account of the developed ontology, starting from its intended application to its final schema as per the NeON methodology. Therefore, the section provides a brief overview of the NeON methodology before delving into the ontology's requirements, specifications, and proposed competency questions. Furthermore, this section highlights the resources that are used in the development process. Finally, the overall ontology schema will be presented in detail.

6.3.1 The NeON Methodology: An Overview

In the first stage of the NeOn methodology, the Ontology Requirement Specification (ORS) must be established [220]. This involves the formulation of competency questions that help to determine the scope of the ontology, as highlighted by Grüninger and Fox in their early work on ontology design and evaluation [336]. ORS uses competency questions to identify several aspects of the ontology, including the purpose of the ontology, its intended users and uses, and the requirements that the ontology should meet [340].

After defining the competency questions and the ORS, the ontology expert can then proceed to explore the relevant knowledge resources that are available to develop the ontology. These resources can be categorised into two types: non-ontological (including glossaries, taxonomies, thesauri, and dictionaries) and ontological resources. The incorporation of ontological resources will lead to a more efficient and cost-effective development process, as well as the creation of a more generalisable semantic framework. Therefore, it is advisable to utilise existing ontologies to represent certain concepts to minimise development time and expenses.

Finally, the chosen resources may require modification to suit the intended use of the new ontology. This involves aligning terminologies and concepts, which may necessitate removing or adding axioms and restructuring the architecture. However, it is essential to ensure overall consistency, which may require an iterative process of revising the model until satisfactory performance has been achieved.

6.3.2 LCA-based Dynamic Environmental Performance of the Building Framework

The intended use of the ontology for an LCA-based dynamic environmental performance assessment is to contextualise and streamline its application in the building domain by incorporating various data sources and modelling techniques to make it more responsive to changing operating conditions, and meet the need to reduce energy consumption and carbon emissions. This ontology aims to give meaning to the different artefacts (e.g., sensory) and processes that are involved in the environmental assessment of buildings, and also describe their possible connections and relationships. In addition, the ontology facilitates the handling of data heterogeneity by introducing a central model to ensure a seamless flow of information. Therefore, based on the intended use of the ontology, there are several requirements that must be considered. First, the ontology should have the ability to describe various concepts, including data types (e.g., sensory data) and different modelling techniques. Second, it must establish links between the LCA domain and related concepts with building objects.

6.3.3 Ontology Requirements Specification

As per the the NeOn methodology, the process of specifying the ontology requirements is an iterative and incremental workflow that begins by identifying the ontology's goals and scope, and concludes by identifying its terminology. This process leads to the creation of an Ontology Requirements Specification Document (ORSD), as shown in Table 6.1.

Table 6.1: Ontology requirements specification document

Purpose of the ontology	The ontology serves as a tool to aid non-LCA experts in implementing control strategies and corrective actions aimed at reducing the environmental impact of buildings. This involves integrating semantic models, dynamic data, ML, and optimisation algorithms to facilitate the decision-making process.
Domain and scope	Environmental assessment, built environment, LCA, building energy management.
Intended users	Facility managers, LCA practitioners
Intended uses	• Exploring options and scenarios with their associated impacts • Monitoring temporal changes • Run energy simulation
Knowledge resources	Existing domain models and methodologies, scientific literature, and domain experts.
Requirements and terminology	See competency questions tables

6.3.4 Competency Questions

Competency questions are of significant importance in the process of developing an ontology because they assist in the identification of key concepts and their interrelationships. Therefore, it is essential for the ontology to provide satisfactory answers to the competency questions. According to [210, 217], competency questions provide an objective measure to evaluate the appropriateness of an ontology's basic structure and the adequacy of its level of detail. In this context, it is important to ensure that an ontology only includes the necessary level of detail to meet the specified requirements, which also highlights the significance of effective requirements elicitation in the ontology's development.

To develop competency questions that accurately reflect the domain of interest, a thorough understanding of the studied domain is essential. This was achieved by identifying the key concepts that constitute the use cases of LCA in the building domain and using UML sequence diagrams to establish the relationship between the identified concepts. It is important to note here that this step serves as a guide to formulate the competency questions and it does not attempt to capture the entirety of the domain knowledge or its complexity.

Figure 6.1 is a crucial tool to identify the key concepts that are required to implement the use case that is investigated throughout this thesis. This figure provides a comprehensive view of the requirements that are necessary to conduct LCA in the built environment domain in general. The development of this figure was informed by the semantisation of use cases technique, which was developed during Stage 2 of the research project and is detailed in Appendix B.

The main concepts include spatial scope, domain, life-cycle stage, scope of the LCA, intended use, enablers, actors, and LCA dynamic elements. Each concept represents a set of entities that can be considered to define and deliver LCA use cases. This figure encompasses numerous entities that are pertinent to the studied domain. However, for the purpose of developing and specifying the scope of the current ontology, entities that are irrelevant or out of scope have been greyed out and are excluded. These entities are recommended for future ontologies that may require a broader scope.

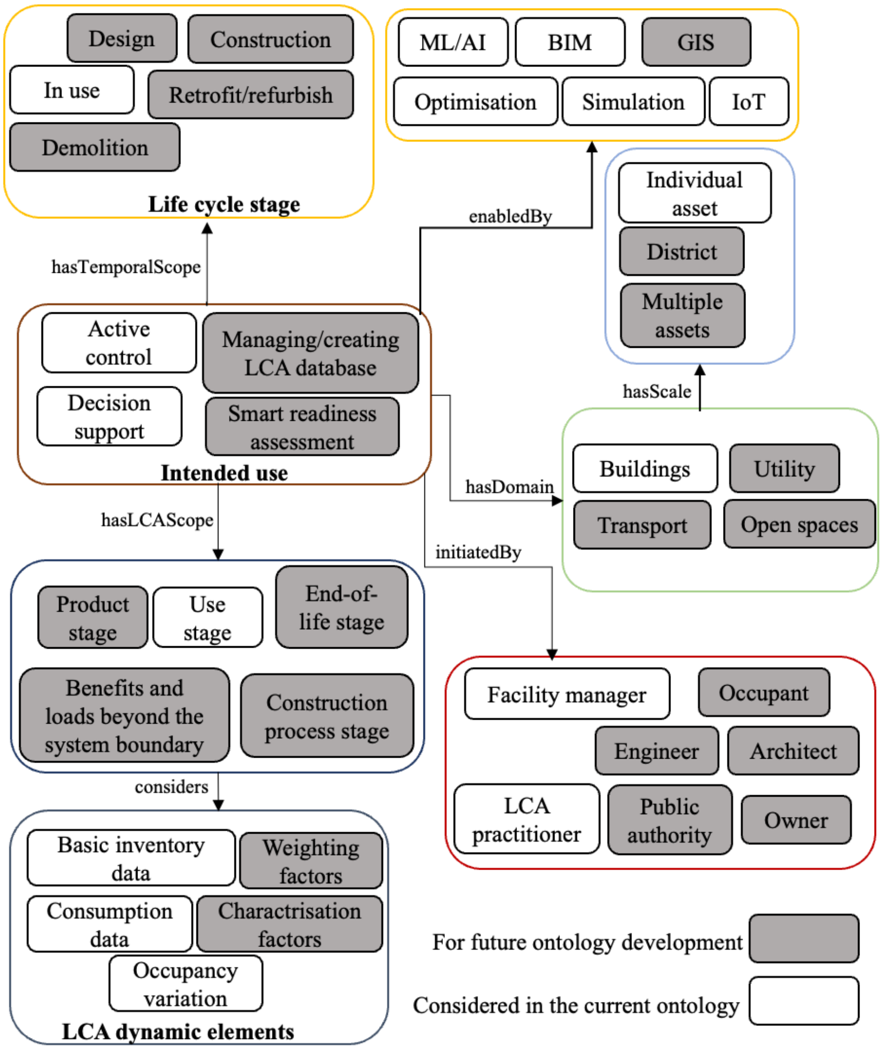


Figure 6.1: Main concepts in the developed ontology

The spatial scope concept specifies the physical boundaries of the asset, while the life-cycle stage identifies the relevant stage in the life cycle of the building. In the current research, a building was identified in the spatial scope and domain because the focus of the use case was a space within one building, and the life-cycle stage was the use phase because the use case focused on energy consumption during this phase. Similarly, the enablers specify the digital resources that are required for the implementation of the use case, such as data collection sources, digital models, and modelling techniques. Entities in the remaining concepts were determined following the same principle.

This step provides a clear picture of the relevant concepts and entities for the current use case. It also highlights the use of several models from different domains, which necessitate the use of semantics to streamline and automate the LCA. However, it is crucial to note that the actors who are involved in the use case do not necessarily have expertise in all of these domains and tools. Hence, there is a need for a system that is capable of integrating these heterogeneous resources and tools. Furthermore, by identifying the relevant concepts and entities that are involved in the use case, this figure serves as a basis for developing a formal representation of the use case using ontology. Finally, Tables 6.2 and 6.3 present two sets of competency questions that were identified as a result of the process.

Table 6.2: Generic competency questions

Number	**Question**
QC1A	What sensors are observing a specific location?
QC2A	What observable properties are being observed by a specific sensor?
QC3A	What controllable parameters are associated with a specific scenario?
QC4A	What power consuming equipment and devices are present in a given space?
QC5A	What simulations are associated with a specific energy profile?
QC6A	Which energy profile is associated with a specific HVAC element?
QC7A	Is there a deviation from the baseline values of energy consumption?

Table 6.3: Competency questions for LCA modelling and scenario description

Number	Question
QC1B	What is the goal/intention/application of the LCA study?
QC2B	What is the functional unit(FU) for a given LCA assessment?
QC3B	What activities are associated with a specific LCI?
QC4B	What is the impact assessment for a specific optimization scenario?
QC5B	Which Life Cycle Impact Assessment (LCIA) method was used?
QC6B	Which impact categories are considered in the LCA model?

6.3.5 Resource Reuse: Ontological Resources

The proposed framework for conducting LCA in the building domain is a multi-disciplinary effort that involves a number of fields of study. In these fields, significant efforts have been made to establish well-defined ontologies that formalise the concepts used within them. Incorporating relevant ontological resources from these fields can be critical to ensuring that the developed ontology is grounded on authoritative sources and is compliant with existing domain-specific knowledge. The NeON methodology advocates the approach of leveraging the rich semantics of established ontologies for efficient ontology development.

6.3.5.1 Sensor Ontology

The LCA-based framework is dependent on gathering data from a number of different sources, including the use of sensor devices. To properly represent the data collection process in the proposed framework, it is essential for the ontology to include an accurate representation of sensors and their readings. In this regard, many efforts have been made in the past decade to design ontologies that capture the abstract concepts of sensors and observations. One noteworthy framework in the literature is the Semantic Sensor Network (SSN) ontology. The SSN-XG, a group within World Wide Web Consortium (W3C), has defined an ontology that specifies the capabilities and properties of sensors, the sensing process, and the resulting observations [341]. The SSN ontology has been selected for this study because it extends the Observation and Measurement framework [342]—which is another ontology that lacks the ability to represent sensor devices and sensing processes—by covering the sensors and their relationships. This is demonstrated in the entity-relationship diagram that is shown in Figure 6.2. The SSN ontology represents sensors as devices that perform a sensing process via an observation capability. The sensing process receives inputs for a stimulus and then produces outputs describing a certain property of a phenomenon (i.e., a feature of interest). Furthermore, to represent the measurements made by sensors in a standardised way, it is essential to use units of measurement that are semantically modelled. This is achieved through the use of the Ontology for Quantities, Units, Dimensions, and Data Types (QUDT), which was developed by NASA [343]. QUDT provides a comprehensive vocabulary of quantities, units, dimensions, and data types that can be used to semantically model the units used in sensor observations.

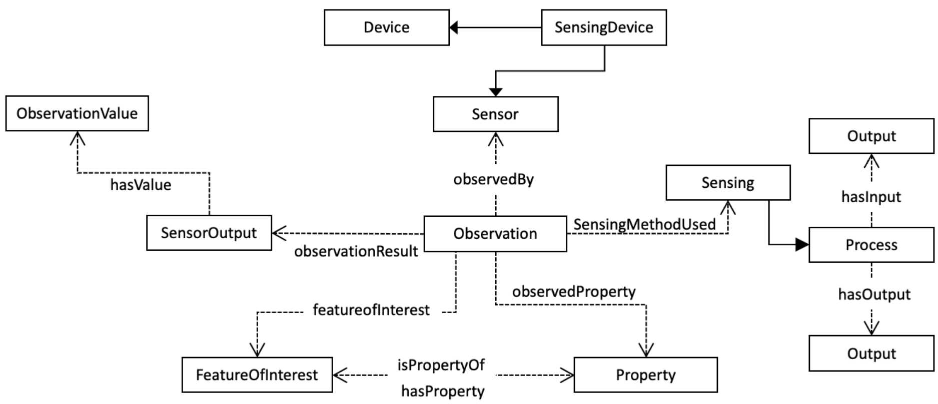


Figure 6.2: Key concepts and relations in the SSN ontology; adapted from [341].

6.3.5.2 Building Ontology

To effectively apply the developed ontology in the building domain, it is crucial for the ontology to capture the relevant semantics of buildings. This requires a comprehensive knowledge domain model that can effectively represent the physical objects within a building that are observed and analysed. Examples of such objects include indoor spaces and technical systems. An ontology for buildings has previously been developed in the form of ifcOWL [344], which provides a semantic model for buildings and their components. Furthermore, ifcOWL is an RDF-based representation of the IFC standard, which serves as a data schema and a file format for exchanging BIM data. In addition, ifcOWL is an extensive ontology that comprises 1,294 classes, 1,573 object properties, and five data properties. An example of the classes that are included in the ifcOWL ontology are IfcBeam, IfcWall, IfcRoof, IfcMaterials, IfcBuilding, and IfcSpace, among others. In the context of the current ontology, ifcOWLSpace can be associated with the SSN feature of interest, which allows for the localisation of sensors.

6.3.6 Resource Reuse: Non-ontological Resources

To enable the reuse of non-ontological resources, classes were derived from the established standards and commonly used terms in the relevant fields. Specifically, for the LCA model, classes were drawn from ISO 14040 and 14044, which are recognised standards in the LCA domain. For other models (e.g., simulation, prediction, and optimisation), classes were identified based on the terms that are frequently employed in the scientific literature and software tools. This approach resulted in the creation of a graph-like structure that consists of interconnected entities.

6.4 Ontology Schema

This section describes the development of an ontology schema to facilitate the ontology's construction process. To achieve this goal, an abstraction of the ontology schema was created, as shown in Figure 6.3, which considers both ontological and non-ontological resources—specifically, the SSN and IfcOWL ontologies, along with relevant standards to identify new entities. The ontology schema is composed of three modules, namely the Service module, the Observation module, and the Building module. This modular design aims to align the entities that exist in the utilised ontological resources with the terms that are identified in the competency questions. A UML sequence diagram to identify further interrelationships between entities in the ontology was developed, which will be discussed next. Furthermore, a detailed representation of each module is given in the following sections to ensure a clear understanding of the ontology's structure and its underlying components.

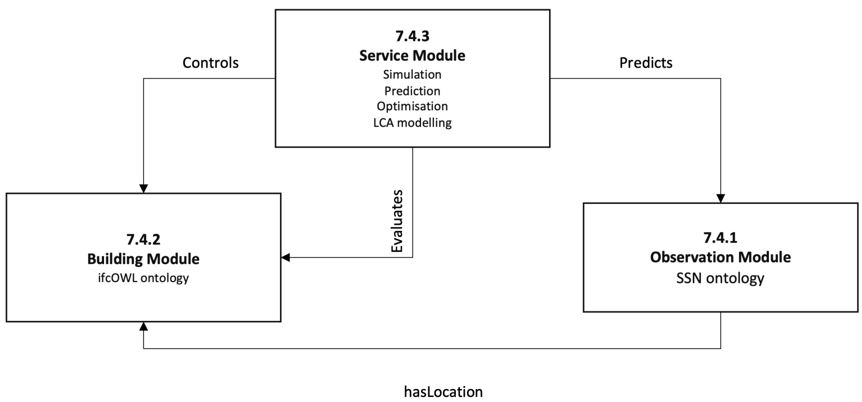


Figure 6.3: Modular schema of the developed ontology

A UML sequence diagram was created to visualise and illustrate the interactions and information exchange between the various tools and techniques that are involved in the implementation of the use case that is investigated in the current research (Figure 6.4). This diagram provides a concise representation of the use case and can be used to illustrate the steps in the process of leveraging sensory data, ML, and optimisation to compare the environmental performance of the operation strategies. This diagram is also relevant to the developed ontology and semantic modelling because it can be used to depict the sequence of events and interactions between the ontology classes, and facilitate interoperability and reuse of data and knowledge.

The sequence diagram illustrates the interaction between the graphical user interface (GUI), semantic middle-ware, and modelling services, which are essential components of the proposed solution. The GUI provides the interface for the LCA practitioner to interact with the system by generating queries, receiving results, and providing feedback. The semantic middle-ware acts as an orchestrator between the user and modelling services, processes the queries, retrieves relevant data, stores the results, and coordinates the execution of tasks and workflows within the system. The modelling services—comprising energy simulation, LCA calculation, optimisation, and prediction—provide the core functionalities of the system. The sequence diagram showcases the step-by-step information exchange and interactions between the system objects, which demonstrates how the use case can be implemented. The LCA practitioner can generate queries to create an LCA for a particular space, optimise and predict specific parameters, or compare scenarios based on their environmental performance.

While the sequence diagram is based on a thorough understanding of the use case and the actual sequence that was followed to generate the results in the previous chapters, it is important to note that the diagram has not been tested in a real system to verify its validity. In other words, the diagram represents an abstract model of the system's behaviour. While it accurately represents the intended interactions between the different components, it has not been implemented and tested in a real-world setting. Therefore, this diagram should be viewed as a

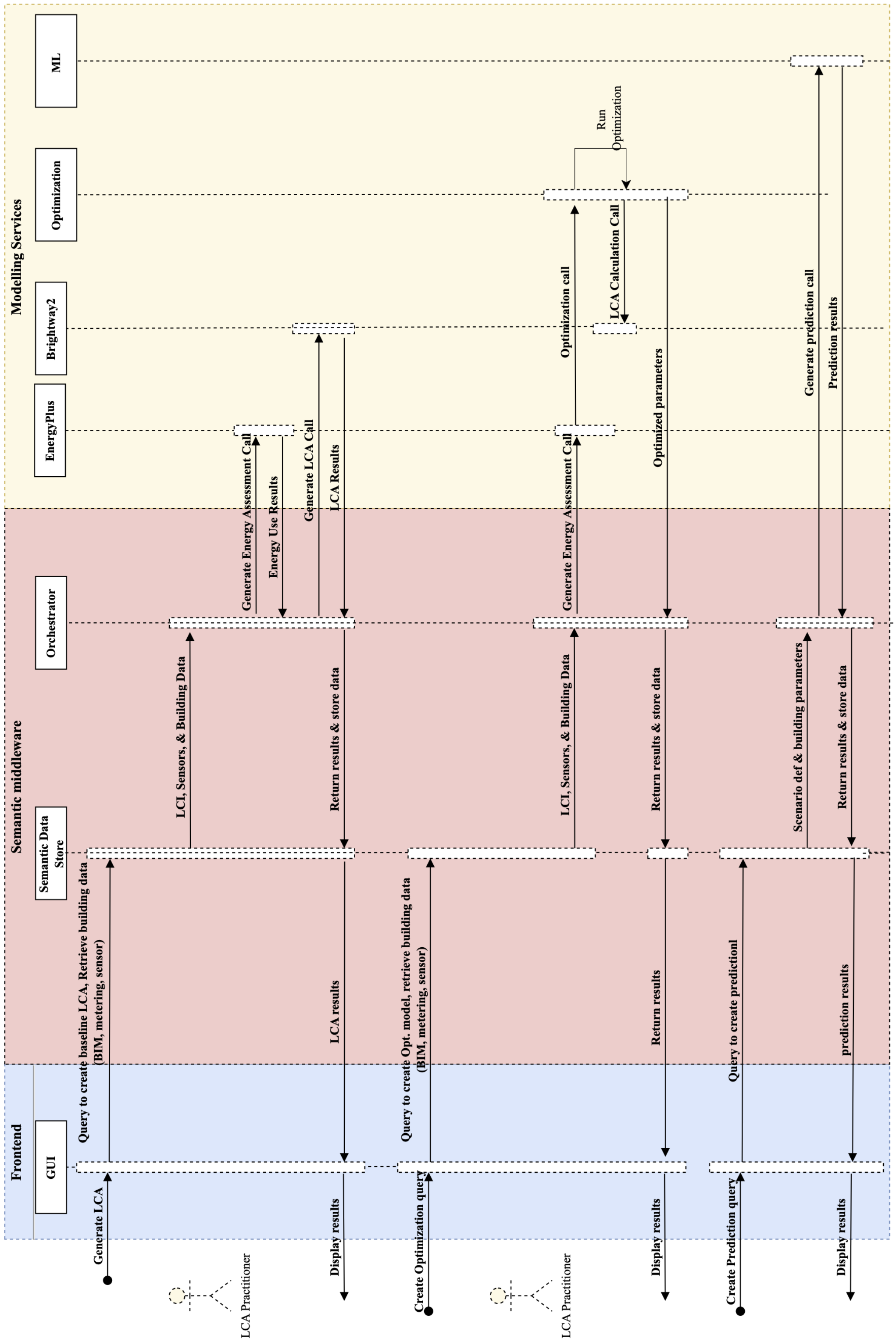


Figure 6.4: UML sequence diagram of the use case

visual representation that can be used to understand the workflow of the system, rather than as a guarantee of its actual behaviour.

6.4.1 Observation Module

The observation module is a crucial component of the developed ontology, as shown in Figure 6.5. This module describes the process of observing a property of an entity through sensors. Through this module, the observable properties of a particular entity can be identified, along with the sensors that observe those properties. Additionally, this module aligns with other modules by identifying relationships with concepts from other modules. For example, the SSN alignment with the building module is established through the 'hasLocation' relationship. Furthermore, the 'Parameter' entity from the Service module is associated with the 'ObservableProperty' class through an 'equivalent' relationship. The 'Parameter' class is an input to the 'Prediction' entity, which refers to potential input parameters to the prediction model. The 'Parameter' entity represents the properties that are observed by the sensors, such as indoor CO_2 and temperature.

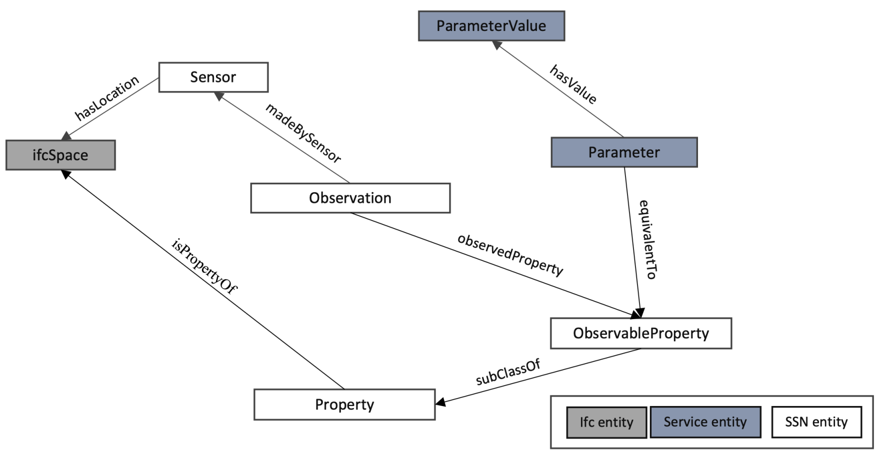


Figure 6.5: Observation module entity relationship diagram

6.4.2 Building Module

Figure 6.6 gives a partial representation of the building's ontology, and primarily aims to depict the relevant ifcOWL classes for the current ontology. The root of this module is an 'ifcObject', which is a superclass that has several subclasses, including 'ifcElements' and 'ifcSpatialStructureElement'. These subclasses are important because they have explicit relationships with the other two modules. As illustrated in the previous module, the location of the sensor is defined through a 'hasLocation' relationship, and the location is an 'ifcSpace'. Additionally, the services module aims to develop multiple models that are applicable to a specific 'ifcElement', based on the work carried out in the previous two chapters (i.e., the ventilation system). The ventilation system is an 'HVACElement' that is based on ifcOWL notation. The linkage between the 'HVACElement' and the service module is established through the 'hasEnergyProfile' relationship, which links the 'HVACElement' to the simulation entity in the service module.

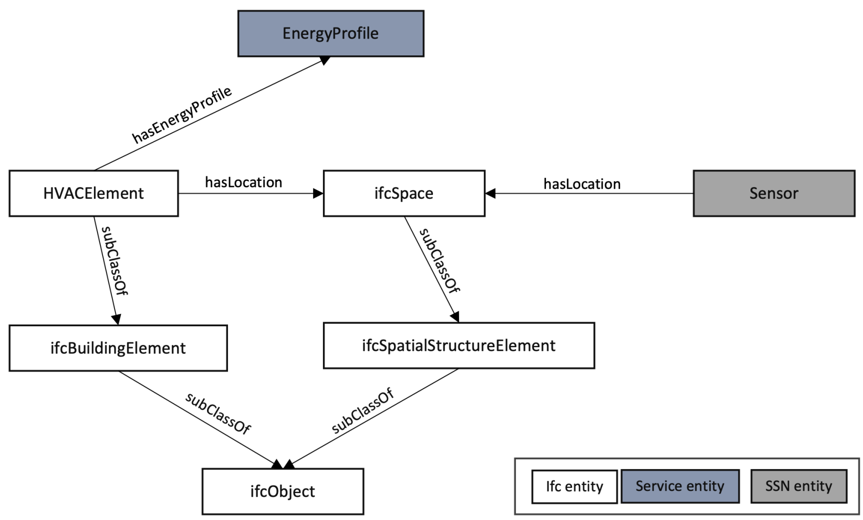


Figure 6.6: Building module entity relationship diagram

6.4.3 Service Module

Figure 6.7 depicts the Service module of the developed ontology, which serves as the backbone of the ontology, hosting the primary framework components that were presented in Chapter 4 (i.e., simulation, prediction, optimisation, and LCA models). The alignment of the service module with the observation and building modules has been discussed in the previous sections. Consequently, the current section focuses on the interplay between entities within the service module. The Simulation class generates energy simulation models for the 'HVACElement'. Similarly, the Prediction class takes some parameters as input and outputs predicted values. The results of both the Prediction and Simulation classes are then used to inform a specific optimisation scenario. This optimisation scenario is implemented via the Optimisation class, which is linked to the Simulation and Prediction classes via 'hasEnergyProfile' and 'usedBy' relationships ,respectively. The results of the Optimisation class (i.e., the optimal values) inform the LCA model via the Demand class. The Demand class has a relationship with the LCI class via the 'hasDemand' relationship. The LCA model includes several classes that cover the most important aspects of the LCA methodology, including the 'LCADatabase', 'ImpactAssessment', 'AssessmentMethod', Activity, and 'ImpactCategory' classes. Other classes were also identified along with their relationships but were not included in the figure because it is only intended for illustrative purposes. A complete representation of the actual ontology will be presented in the next section, where the ontology implementation in Protégé is discussed.

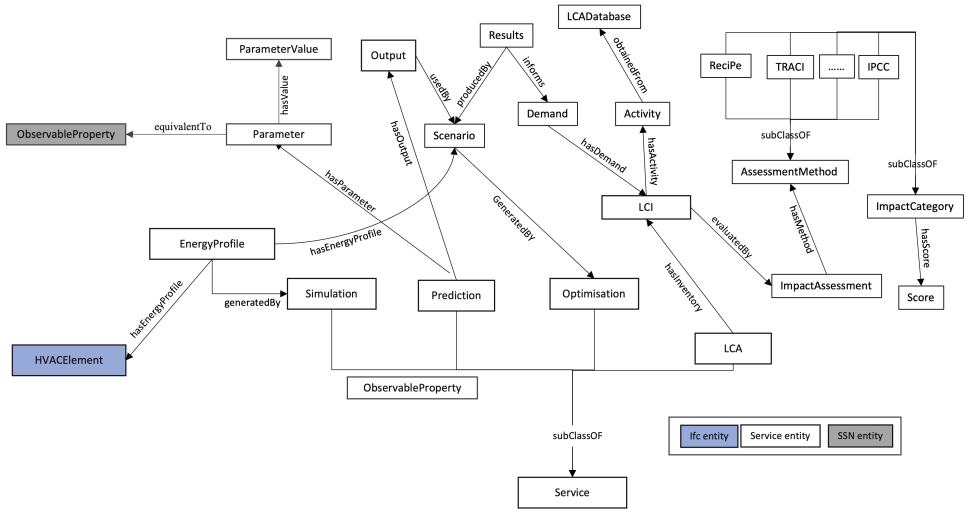


Figure 6.7: Service module entity relationship diagram

6.5 Ontology: Implementation and Evaluation

The ontology has been developed using Protégé, which is a widely used open-source ontology editor and knowledge management tool. Running a reasoner is an essential step in the ontology development because it ensures the logical consistency and coherence of the ontology. The HermiT reasoner was used to verify the consistency and classify the concepts within the ontology. The proposed ontology consists of 135 logical axioms, 35 classes, and 25 object properties. These metrics provide insight into the structure of the ontology and can aid in assessing the usability of the developed ontology. Figure 6.8 presents the ontology. Figure 6.9 presents the main classes, object properties, and instances.

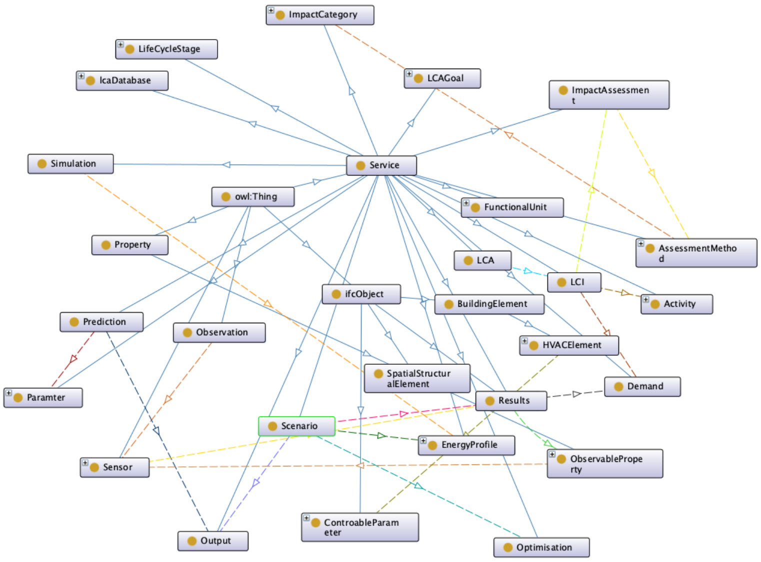


Figure 6.8: Ontology representation in Protégé

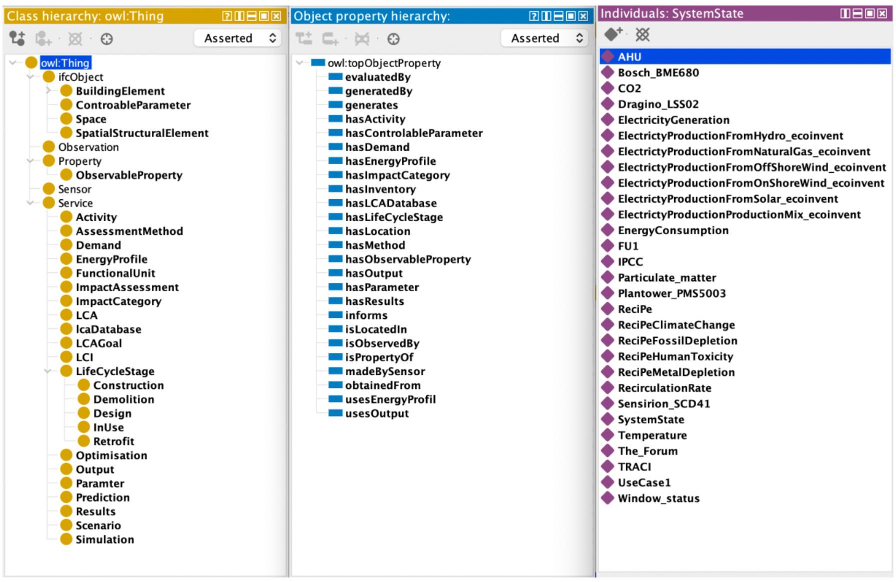


Figure 6.9: Main classes, relationships, and instances of the ontology

To evaluate the developed ontology, it is necessary to refer to the identified competency questions and then determine whether the ontology is able to provide satisfactory answers to these questions. To achieve this aim, SPARQL queries (a commonly used query language for querying RDF data) were used on the developed ontology. The queries were run using the Fuseki server, which is an open-source SPARQL server that allows RDF data to be stored and queried. Furthermore, to ensure that the ontology is not only logically consistent but also practically useful in real-world scenarios, it was instantiated with example data from the developed use case in Chapter 4. Evaluation queries were then developed for some of the identified competency questions. These queries are presented in Figure 6.10 to 6.15, which illustrate the validity of the developed ontology.

In Figure 6.10, the user seeks information about the available sensors within a specific space, in this case the *Forum*. The query returns the names of the sensors and their location. In the next query, the user can retrieve the available sensors and their observable properties (Figure 6.11). In Figure 6.12, the user is interested in the controllable parameters of the HVAC system. The query returns two values: system state and recirculation rate. To answer the second set of competency questions related to the LCA model, Figures 6.13 and 6.14 demonstrate how the user can query information about the goal and functional unit of the LCA model. The final query in Figure 6.15 identifies the activity that is involved in the LCA study and its associated unit processes from the LCA database.

The test queries were based on a subset of identified competency questions and demonstrate the effectiveness of the developed ontology in providing answers to questions from various domains. The ontology's ability to offer satisfactory responses to queries related to different domains underscores the significant advantages of leveraging semantics to integrate information and data from heterogeneous sources, which is a critical step towards enhancing interoperability and information exchange across several domains.

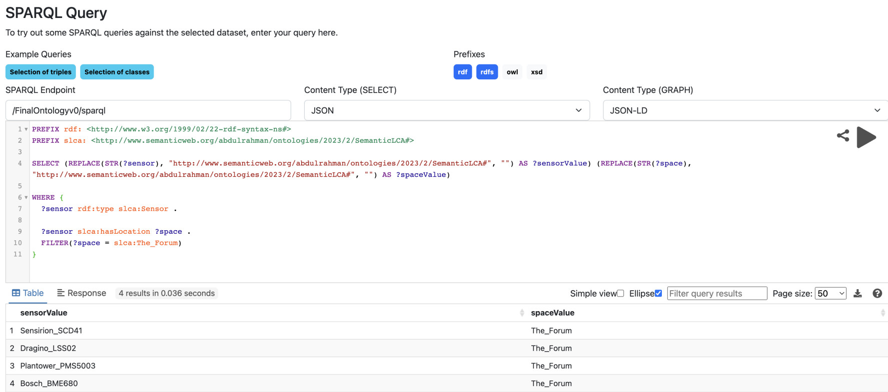


Figure 6.10: SPARQL query for QC1A

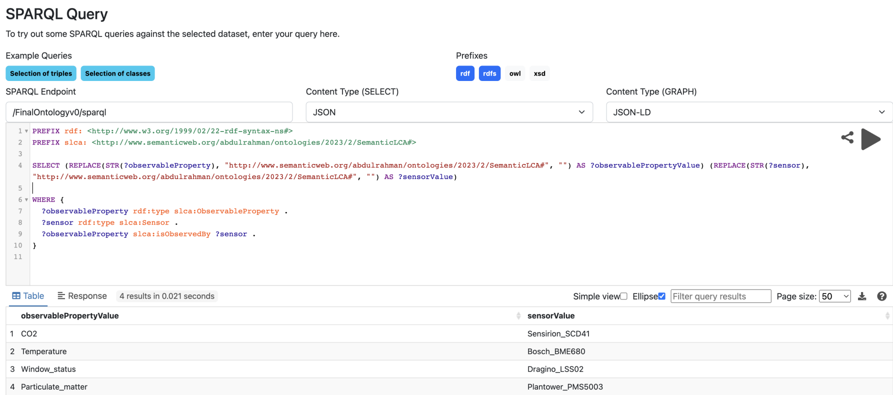


Figure 6.11: SPARQL query for QC2A

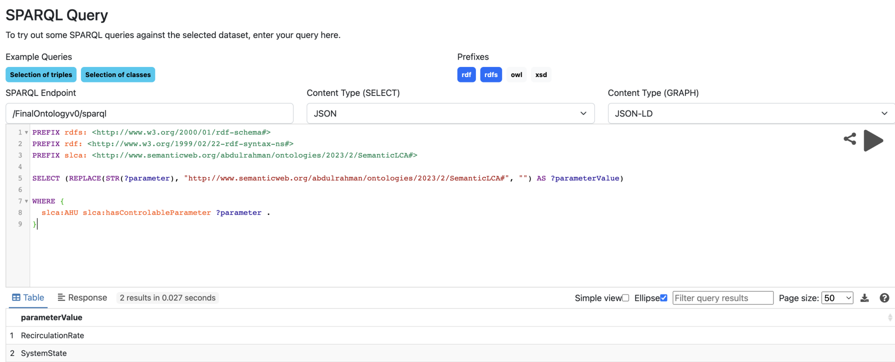


Figure 6.12: SPARQL query for QC3A

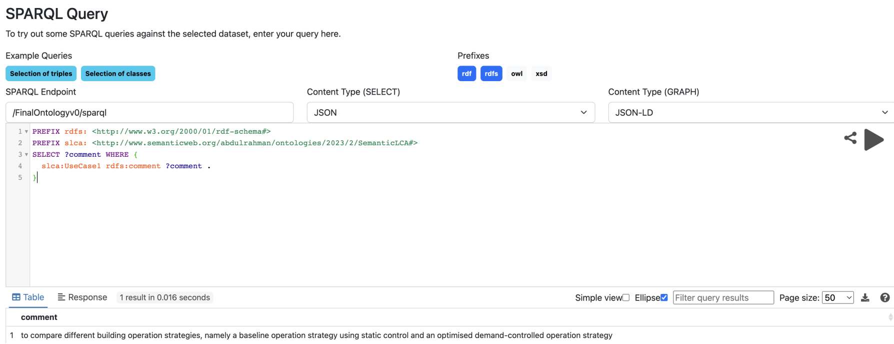


Figure 6.13: SPARQL query for QC1B

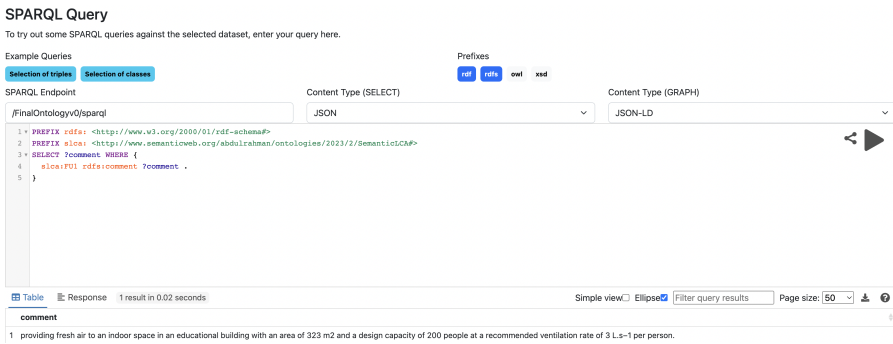


Figure 6.14: SPARQL query for QC2B

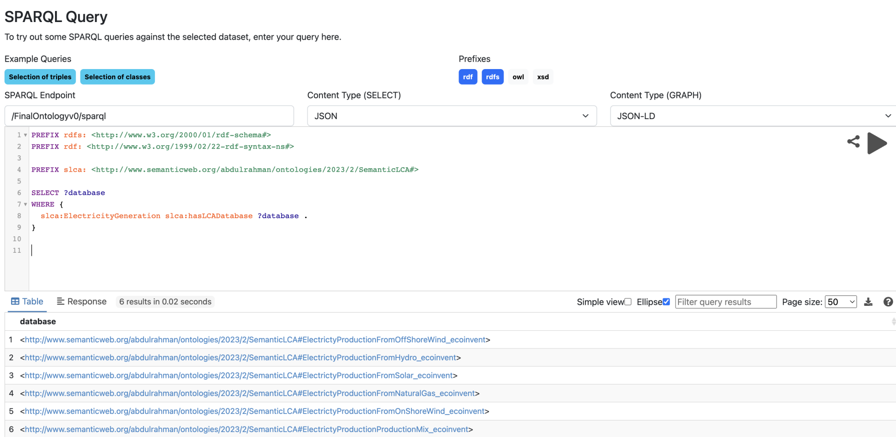


Figure 6.15: SPARQL query for QC3B

6.6 Summary

This chapter aimed to investigate how the application of a semantically-enabled LCA can streamline and automate the process of assessing the environmental impact of buildings. The research question asks,

Can a semantic web approach provide a sound basis to facilitate and streamline the applications of LCA in buildings?

To answer this question, the proposed solution was developed based on the understanding of the use case carried out in this thesis and the challenges that were identified from the literature review. This chapter demonstrated the importance of semantic technologies in improving the effectiveness and efficiency of a LCA application in the building domain. This chapter discussed the steps that were taken to develop the proposed solution, including the conceptualisation of the use case and the development of a lightweight ontology.

To develop the ontology, a methodical approach was followed which identified the domain concepts, defined their properties, created relationships between the concepts, and reused existing ontological and non-ontological resources. Competency questions were also identified to ensure that the ontology can provide a satisfactory performance. In the ontology's development stage, an ontology schema was developed by creating three interconnected modules, namely: the Observation module, Service module, and Building module. These modules were created based on the requirements that were identified in the ontology requirements specification stage. The Observation module defines classes and properties related to sensor observations. The Service module includes classes and properties related to several models, which are required to improve the environmental performance of buildings . Finally, the Building module contains the classes and properties that are related to the building itself.

The ontology was then evaluated by running SPARQL queries on the developed ontology using the Fuseki server. The ontology was instantiated with example data from the developed use case in Chapter 4, which aimed to ensure that it was not only logically consistent but also practically useful in real-world scenarios. Evaluation queries were then developed for some of the identified competency questions. The results demonstrate the effectiveness of the developed ontology in providing answers to questions from various domains. The ontology's ability to offer satisfactory responses to queries related to different domains underscores the significant advantages of leveraging semantics to integrate information and data from heterogeneous sources. The integration of this information and data is a critical step towards enhancing interoperability and information exchange across several domains.

In summary, this chapter has emphasised the importance of considering existing relevant ontologies to develop a robust system that is capable of extracting meaning and integrating several models. Furthermore, this chapter provides insights into how semantic technologies can facilitate and streamline the application of LCA. Finally, the work in this chapter can also serve as a starting point for further research and development in the field of semantically-enabled LCA for buildings.

Chapter 7

Conclusion

This chapter will provide a comprehensive reflection on the findings of the present research. To achieve this, the central hypothesis and research questions that were posed in Chapter 1 will be revisited, and each research question will be addressed based on the observations and findings that were presented in the related chapters. Additionally, the key contributions of the present study to the body of knowledge will be identified. Finally, the limitations of the study will be discussed and a number of recommendations will be made for future research, aimed at building on these limitations and improving this research.

7.1 Research Findings

This section will present the research findings, with a particular emphasis on addressing the central hypothesis of the study, which was formulated as follows: "*A semantic-based approach can facilitate the process of LCA and improve the accuracy of the LCA results by leveraging the value of dynamic data, learning systems, and digital built environment resources.*"

As previously stated in Chapter 1, the hypothesis was formulated into four research questions, each of which will be thoroughly examined in this section based on the findings presented in the related chapters. Finally, a summary discussion of the research hypothesis will be provided.

7.1.1 LCA Application in the Building Domain

The first research question asked:

What are the key limitations of current LCA methods that affect the accuracy and widespread adoption of LCA in the building domain?

The first research question was the primary focus of the exploratory stage of the research, which aimed to investigate the current applications of LCA in buildings. Therefore, a comprehensive literature review was conducted, which facilitated a thorough analysis of the state-of-the-art research in the field of LCA applied to buildings. The findings of the literature review not only identified the shortcomings of the proposed solutions but also recognised the requirements and functionality that are required to streamline LCA throughout the life cycles of the asset. Moreover, this stage provided a foundation for the modelling choices and the development of the proposed framework.

The literature review in Chapter 2 highlighted several limitations and gaps in the current LCA solutions. These include limitations in the semantic information and interoperability of current software solutions in BIM-LCA integration. Although the complexity and dynamic nature of building LCA requires a comprehensive approach to explore various scenarios, the scope and capabilities of the existing decision support tools suffer from a number of limitations. In particular, limited research exists on the application of LCA at the district and city-wide level. There is also a need to understand the level of detail required at the building level and informative attributes at the district level to deliver reliable and sound LCA results. Furthermore, there is a lack of consideration for temporal information in LCA studies for buildings. Few studies have explored the use of dynamic data, such as IoT devices, for indoor environmental measurements, which highlights the need for further research in this area. Finally, the outcomes of this stage informed the development of an overarching framework that leverages dynamic data and learning systems, and also explores the role of semantics in the LCA applications for buildings.

7.1.2 Optimising the Environmental Performance of Buildings

The second and third research questions asked:

Can access to dynamic data provide more accurate accounts of the environmental impacts during the operation stage?

How can machine learning and optimisation be leveraged to reduce the environmental impact of buildings?

These research questions were addressed separately in Chapters 4 and 5. However, the integrated nature of the proposed framework requires a simultaneous discussion to achieve a comprehensive understanding of how dynamic data, ML, and optimisation can enhance the energy and environmental performance of buildings. Therefore, at this point, a different approach was taken to address both questions together. Therefore, the third stage of the research methodology represents the culmination of the previous two stages and serves as the core contribution of this study. In this stage, a specific use case was selected to demonstrate how the generic framework that was developed during the participation with the SemanticLCA project could be applied to address the second and third research questions. Based on the literature review and engagement in the SemanticLCA research project, it was found that the energy performance gap and the lack of consideration for dynamic factors were major challenges in reducing the environmental impact of a building's energy consumption during the operation phase. To address these challenges, the proposed framework leverages the concepts of semantic interoperability and dynamic data, and utilises ML and optimisation algorithms to develop a decision support system that helps to contextualise and translate information into actionable measures. By integrating various domain models and data sources, the methodology aims to provide a holistic approach that takes multiple (often conflicting) objectives into account during the operation of a building. Thus, the work that was conducted in the third stage demonstrates the practical application of the developed framework in addressing the research questions.

A use case was developed with the intention of minimising the environmental impacts of a building's energy consumption through an optimised mechanical ventilation system strategy that considers dynamic indoor conditions (e.g., CO_2) that are captured by indoor sensors using ML and an optimisation technique (i.e., the genetic algorithm). This was compared to a baseline scenario that represents a schedule-based, static operation strategy, which was modelled using energyPlus.

This study employed ML in two ways. First, ML models were developed to predict CO_2 levels in the case study site, based on different sets of input variables. These models were developed to predict CO_2 levels at two different time horizons, 1 hour and 24-hour ahead. Two approaches were employed for the 1-step prediction: univariate models relying only on CO_2 lags, and multivariate models that employed other parameters (e.g., temperature and humidity etc.). These models represent a crucial input to the optimisation strategy because they provide information about the indoor conditions for the next 24 hours, which is a key factor in identifying the optimal operation strategy for the mechanical ventilation. Second, an ANN model was developed to predict both energy consumption and indoor CO_2 levels based on the output of energy simulations and a different set of input variables. This model was then used by the GA as a fitness evaluation to obtain the optimal operation strategy.

An optimisation strategy for mechanical ventilation systems using GA was developed. The objective here was to minimise energy consumption while ensuring that indoor CO_2 levels remain within acceptable limits. This strategy employed a ML model that was trained on EnergyPlus simulations to predict energy consumption and CO_2 levels, which was then used as the fitness evaluation function for the genetic algorithm (as previously mentioned). The optimisation process adjusts the recirculation rate, which reduces the required heating energy and the required outside air volume, thereby minimising energy consumption. The optimisation strategy that was proposed in this study was able to reduce energy consumption by 35% when compared to the baseline scenario, while also ensuring that indoor CO_2 levels remain below 900 ppm.

Finally, the LCA results of the two scenarios, baseline and optimised, reveal that the latter has lower environmental impacts across all impact categories, especially in terms of climate change. Furthermore, the results highlight the importance of optimising a building's energy performance to reduce its impact on the environment, even when transitioning towards cleaner and renewable energy technologies.

7.1.3 The Role of Semantics in LCA

The final research question asked:

Can a semantic web approach provide a sound basis to facilitate and streamline the applications of LCA in buildings?

Chapter 6 used a solution-based approach to address the final research question, which developed a lightweight ontology using a methodical approach. The methodology identified domain concepts, created relationships between them, and reused existing resources. The ontology was instantiated with example data from the use case to ensure logical consistency and practicality. In addition, evaluation queries were tested and the results demonstrate the effectiveness of this approach to provide answers to a number of domain questions.

The modular approach that was used in developing the ontology schema can enable both scalability and extensibility. The current ontology schema consists of three interconnected modules (i.e., the Observation, Service, and Building modules) but other modules can be incorporated for scalability purposes. For instance, a District module can be added to represent urban objects such as building blocks. The ontology can also be extended with more concepts based on a set of requirements that are identified in the ontology requirements specification step. The inclusion of relevant concepts in the current ontology was guided by the need to identify the scope and requirements of specific use cases. Although new concepts can be added to the current modules, they must align with the scope and requirements of any extension to the current ontology.

Finally, this work emphasised the importance of utilising semantics to improve interoperability and information exchange across various domains. Therefore, it is hoped that this study may serve as an example for further research and development in the field of LCA for buildings using semantics.

7.1.4 Revisiting the Hypothesis

The central research hypothesis of this study can be evaluated based on the previous discussion of the four research questions. The use of dynamic data, particularly data from sensors, allows for a better understanding of the operating conditions, as opposed to relying on assumptions regarding the dynamic building conditions, which could lead to the overestimation or underestimation of the true environmental performance of buildings. The incorporation of dynamic data captured by sensors has resulted in more reliable data that inform operational decisions. Moreover, the use of ML and optimisation, while leveraging real-time data, has improved the reliability of the models, and their prediction and proposed operation strategies. This resulted in more accurate LCA calculations in the building domain. Finally, this study developed an ontology and incorporated several semantic models, including SSN and ifcOWL, and other data sources. The developed ontology facilitated the alignment of these resources and streamlined the application of LCA in the building domain. This successful implementation and utilisation of dynamic data, ML, and optimisation models, along with the developed ontology, provide evidence in support of the research hypothesis.

7.2 Contributions

The contributions resulting from this thesis will be described in this section. These contributions can be attributed to the development of an integrated framework to optimise building energy and environmental performance, which was discussed in Chapters 4 and 5, as well as the semantic modelling of LCA in the building domain, which was covered in Chapter 6.

- The first main contribution to the body of knowledge results from the integrated framework, which can be divided into two minor contributions.
  - The first minor contribution relates to the development and testing of two ML models, namely RF and ANN, to predict indoor CO_2 levels for different time steps, specifically one hour ahead and 24 hours ahead. These models consider indoor and outdoor conditions, as well as time-related parameters, to improve the accuracy of their predictions.
  - The second minor contribution relates to the development of an optimisation strategy to control the ventilation systems in buildings, with the goal of minimising energy consumption while maintaining indoor conditions within acceptable limits, particularly the CO_2 levels. These two contributions inform the LCA model, which is carried out based on representative and accurate data that reflect the real operating conditions of the building, rather than relying on average data or broad assumptions regarding the building's dynamics.

- The second main contribution of this thesis was the development of a lightweight ontology for LCA applied to buildings. The modular approach that was taken in developing the ontology schema allows for scalability and extensibility, which enables the incorporation of additional modules as needed. The developed ontology exemplified the potential to improve interoperability and information exchange across different domains by facilitating semantic modelling.

7.3 Limitations and Future Work

The first clear limitation of the present study is that the validation of the proposed framework and the corrective measures, informed by the optimisation results, were mainly based on simulation data. The practical implications of the optimisation strategy were not tested on the actual case study site because the necessary utility meter data and access to the building management system were unavailable. Therefore, the true impact of the optimisation strategy can only be accurately captured through the real-time implementation of the proposed measures in the system. Furthermore, reliance on simulation data may introduce uncertainties and potential discrepancies between the simulated and actual performance of the system, thereby limiting the validity of the proposed framework. To overcome this limitation of the present study, future research should focus on implementing the proposed optimisation strategy in real-time on the case study site, using actual data from the utility meters and the building's management systems. This would enable accurate measurement of the impact of the proposed measures on energy use and indoor conditions, and provide a more realistic understanding of the practical implications of the strategy.

The second limitation of the current study is the narrow scope of the LCA model, which only considered energy consumption of one month during the use phase, which precluded a holistic understanding of the building's energy performance because seasonal and usage variations occur throughout the year. Furthermore, while this study demonstrated the potential benefits of smart systems in reducing the energy use and environmental impacts of buildings, certain critical aspects were excluded, including the environmental impacts of monitoring system hardware (e.g., sensors, gateways, and servers). Excluding these limitations from the studied system may lead to a different conclusion regarding the actual environmental benefits of smart systems. Therefore, in future research, it is crucial to address the limitations of the current study and consider the minimum level of instrumentation that is required to monitor and control the building's systems. This will ensure that the goals of reducing the environmental impacts of buildings are achieved instead of adding to their environmental burdens.

The final limitation of this work is that the developed ontology was designed with a limited scope and scale, and was only intended for the specific use case that was investigated in this thesis. While the ontology proved valuable for the present work, certain aspects should be considered for future research. First, the developed ontology was not integrated into a functional system that is capable of integrating different models and databases through the use of semantics and the developed ontology to store these artefacts in a semantic data store, which would allow the users to write queries to evaluate different scenarios. Moreover, neither scalability nor consideration of a wider range of applications were included in the current version of the ontology. However, the modular approach that is employed in the ontology's design allows for further extensions and the inclusion of other modules in future research. To address this limitation, future work should focus on expanding the scope of the ontology and should consider scalability to support a wider range of applications.

7.4 Closing Remarks

This thesis has encompassed three distinct aspects. The first aspect is the use of semantics to contextualise, integrate, and streamline the information exchange between various components, thereby enhancing the interoperability of LCA. The second aspect is the consideration of dynamic data (e.g., energy use, indoor and outdoor conditions), which enables a more accurate assessment of the environmental impacts. The third aspect is the integration of learning systems, including ML and optimisation, which aid in the decision-making process by exploring different scenarios and identifying the corrective actions that result in the least environmental impacts. It is important to note that despite the progress that has been made in this thesis, there is still a great deal of work to be done to achieve a true semantic model of LCA. Therefore, the work that was conducted in this thesis can be seen as a proof of concept and as part of ongoing research towards a sustainable future.

Bibliography

[1] United Nations, Department of Economic and Social Affairs, Population Division. World urbanization prospects: The 2018 revision. Technical report, United Nations, New York, 2019.

[2] International Energy Agency. World energy outlook 2018. Technical report, IEA, Paris, 2018.

[3] Véronique Monier, Mathieu Hestin, Manuel Trarieux, Shame Mimid, Lena Domröse, Mike Van Acoleyen, Peter Hjerp, and Shailendra Mudgal. Study on the management of construction and demolition waste in the EU. Contract 07.0303/2009/540863/SER/G2, Final report for the European Commission (DG Environment), 2011.

[4] IPCC. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. In T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, editors, *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[5] Chirjiv Kaur Anand and Ben Amor. Recent developments, future challenges and new research directions in lca of buildings: A critical review. *Renewable and Sustainable Energy Reviews*, 67:408–416, 2017.

[6] Cristiane Bueno, Michael Zwicky Hauschild, João Adriano Rossignolo, Aldo Roberto Ometto, and Natália Crespo Mendes. Sensitivity analysis of the use of life cycle impact assessment methods: A case study on building materials. *Journal of Cleaner Production*, 112:2208–2220, 2016.

[7] Christofer Skaar and Rikke B Jørgensen. Integrating human health impact from indoor emissions into an lca: A case study evaluating the significance of the use stage. *The International Journal of Life Cycle Assessment*, 18 (3):636–646, 2013.

[8] Koji Negishi, Ligia Tiruta-Barna, Nicoleta Schiopu, Alexandra Lebert, and Jacques Chevalier. An operational methodology for applying dynamic life cycle assessment to buildings. *Building and Environment*, 144:611–621, 2018.

[9] Bin He, Qijun Pan, and Zhongqiang Deng. Product carbon footprint for product life cycle under uncertainty. *Journal of Cleaner Production*, 187: 459–472, 2018.

[10] Edgar G Hertwich, Saleem Ali, Luca Ciacci, Tomer Fishman, Niko Heeren, Eric Masanet, Farnaz Nojavan Asghari, Elsa Olivetti, Stefan Pauliuk, Qingshi Tu, et al. Material efficiency strategies to reducing greenhouse gas emissions associated with buildings, vehicles, and electronics—a review. *Environmental Research Letters*, 14(4):043004, 2019.

[11] Xiao-Juan Li and Yan-dan Zheng. Using lca to research carbon footprint for precast concrete piles during the building construction stage: A china study. *Journal of Cleaner Production*, 245:118754, 2020.

[12] Catherine De Wolf, Endrit Hoxha, and Corentin Fivet. Comparison of environmental assessment methods when reusing building components: A case study. *Sustainable Cities and Society*, 61:102322, 2020.

[13] N Futas, K Rajput, and R Schiano-Phan. Cradle to cradle and whole-life carbon assessment: Barriers and opportunities towards a circular economic building sector. In *IOP Conference Series: Earth and Environmental Science*, volume 225, page 012036. IOP Publishing, 2019.

[14] Charles Thibodeau, Alain Bataille, and Marion Sié. Building rehabilitation life cycle assessment methodology: State of the art. *Renewable and Sustainable Energy Reviews*, 103:408–422, 2019.

[15] Sahar Mirzaie, Mihaela Thuring, and Karen Allacker. End-of-life modelling of buildings to support more informed decisions towards achieving circular

economy targets. *The International Journal of Life Cycle Assessment*, 25 (11):2122–2139, 2020.

[16] Sergio García-Pérez, Jorge Sierra-Pérez, and Jesús Boschmonart-Rives. Environmental assessment at the urban level combining lca-gis methodologies: A case study of energy retrofits in the barcelona metropolitan area. *Building and Environment*, 134:191–204, 2018.

[17] Bernardette Soust-Verdaguer, Carmen Llatas, and Antonio García-Martínez. Critical review of bim-based lca method to buildings. *Energy and Buildings*, 136:110–120, 2017.

[18] Giuseppe Cardellini, Christopher L Mutel, Estelle Vial, and Bart Muys. Temporalis, a generic method and tool for dynamic life cycle assessment. *Science of the Total Environment*, 645:585–595, 2018.

[19] Ligia Tiruta-Barna, Yoann Pigné, Tomás Navarrete Gutiérrez, and Enrico Benetto. Framework and computational tool for the consideration of time dependency in life cycle inventory: proof of concept. *Journal of Cleaner Production*, 116:198–206, 2016.

[20] Jennifer Andrews. Greenhouse gas emissions inventory reports: Fy 14 briefing. 2014.

[21] Muhammad Asif. An empirical study on life cycle assessment of double-glazed aluminium-clad timber windows. *International Journal of Building Pathology and Adaptation*, 2019.

[22] Youssef O Elkhayat, Mona G Ibrahim, Koji Tokimatsu, and Ahmed AbdelMonteleb M Ali. A comparative life cycle assessment of three high-performance glazing systems for office buildings in a hot desert climate zone. *Clean Technologies and Environmental Policy*, 22(7):1499–1515, 2020.

[23] Yuanli Lyu and Tin-tai Chow. Economic, energy and environmental life cycle assessment of a liquid flow window in different climates. *Building Simulation*, 13:837–848, 2020.

[24] Michael Budig, Oliver Heckmann, Markus Hudert, Amanda Qi Boon Ng, Zack Xuereb Conti, and Clement Jun Hao Lork. Computational screening-

lca tools for early design stages. *International Journal of Architectural Computing*, 19(1):6–22, 2021.

[25] Ruochen Zeng, Abdol Chini, and Robert Ries. Innovative design for sustainability: Integrating embodied impacts and costs during the early design phase. *Engineering, Construction and Architectural Management*, 28(3): 747–764, 2020.

[26] Esmaeel Asadi, Abdullahi M Salman, and Yue Li. Multi-criteria decision-making for seismic resilience and sustainability assessment of diagrid buildings. *Engineering Structures*, 191:229–246, 2019.

[27] Vaclav Hasik, Maximilian Ororbia, Gordon P Warn, and Melissa M Bilec. Whole building life cycle environmental impacts and costs: A sensitivity study of design and service decisions. *Building and Environment*, 163: 106316, 2019.

[28] Inkwan Paik and Seunguk Na. Comparison of environmental impact of three different slab systems for life cycle assessment of a commercial building in south korea. *Applied Sciences*, 10(20):7278, 2020.

[29] Jagat Kumar Shrestha. Assessment of energy demand and greenhouse gas emissions in low rise building systems: Case study of five building systems built after the gorkha earthquake in nepal. *Journal of Building Engineering*, 34:101831, 2021.

[30] Robert Phillips, Luke Troup, David Fannon, and Matthew J Eckelman. Triple bottom line sustainability assessment of window-to-wall ratio in us office buildings. *Building and Environment*, 182:107057, 2020.

[31] James Helal, André Stephan, and Robert H Crawford. The influence of structural design methods on the embodied greenhouse gas emissions of structural systems for tall buildings. 24:650–665, 2020.

[32] Zhixing Luo and Yiqing Lu. Multi-case study on the carbon emissions of the ecological dwellings in cold regions of china over the whole life cycle. *Energy Exploration & Exploitation*, 38(5):1998–2018, 2020.

[33] Benedek Kiss and Zsuzsa Szalay. Modular approach to multi-objective environmental optimization of buildings. *Automation in Construction*, 111: 103044, 2020.

[34] Mattia Manni, Gabriele Lobaccaro, Nicola Lolli, and Rolf Andre Bohne. Parametric design to maximize solar irradiation and minimize the embodied ghg emissions for a zeb in nordic and mediterranean climate zones. *Energies*, 13(18):4981, 2020.

[35] Ran Wang, Shilei Lu, Wei Feng, Xue Zhai, and Xinhua Li. Sustainable framework for buildings in cold regions of china considering life cycle cost and environmental impact as well as thermal comfort. *Energy Reports*, 6: 3036–3050, 2020.

[36] Mateusz Płoszaj-Mazurek, Elżbieta Ryńska, and Magdalena Grochulska-Salak. Methods to optimize carbon footprint of buildings in regenerative architectural design with the use of machine learning, convolutional neural network, and parametric design. *Energies*, 13(20):5289, 2020.

[37] Madjid Tavana, Mohammad Izadikhah, Reza Farzipoor Saen, and Ramin Zare. An integrated data envelopment analysis and life cycle assessment method for performance measurement in green construction management. *Environmental Science and Pollution Research*, 28(1):664–682, 2021.

[38] Seyed Amirhosain Sharif and Amin Hammad. Developing surrogate ann for selecting near-optimal building energy renovation methods considering energy consumption, LCC and LCA. *Journal of Building Engineering*, 25: 100790, 2019.

[39] Camilla Ernst Andersen, Pernille Ohms, Freja Nygaard Rasmussen, Harpa Birgisdóttir, Morten Birkved, Michael Hauschild, and Morten Ryberg. Assessment of absolute environmental sustainability in the built environment. *Building and Environment*, 171:106633, 2020.

[40] Chanjief Chandrakumar, Sarah J McLaren, David Dowdell, and Roman Jaques. A science-based approach to setting climate targets for buildings: The case of a new zealand detached house. *Building and Environment*, 169: 106560, 2020.

[41] Joanna Rucińska, Anna Komerska, and Jerzy Kwiatkowski. Preliminary study on the gwp benchmark of office buildings in poland using the lca approach. *Energies*, 13(13):3298, 2020.

[42] Freja Nygaard Rasmussen, Sara Ganassali, Regitze Kjær Zimmermann, Monica Lavagna, Andrea Campioli, and Harpa Birgisdóttir. Lca benchmarks for residential buildings in northern italy and denmark: Learnings from comparing two different contexts. *Building Research & Information*, 47(7):833–849, 2019.

[43] Patricia Schneider-Marin and Werner Lang. Environmental costs of buildings: Monetary valuation of ecological indicators for the building industry. *The International Journal of Life Cycle Assessment*, 25(9):1637–1659, 2020.

[44] Leonora Charlotte Malabi Eberhardt, Anne van Stijn, Freja Nygaard Rasmussen, Morten Birkved, and Harpa Birgisdottir. Development of a life cycle assessment allocation approach for circular economy in the built environment. *Sustainability*, 12(22):9579, 2020.

[45] Joint Research Centre, Institute for Environment, and Sustainability. *International Reference Life Cycle Data System (ILCD) handbook : review schemes for Life Cycle Assessment*. Publications Office, 2011. doi: doi/10.2788/39791.

[46] Xiaocun Zhang, Kaihua Liu, and Zihua Zhang. Life cycle carbon emissions of two residential buildings in china: Comparison and uncertainty analysis of different assessment methods. *Journal of Cleaner Production*, 266:122037, 2020.

[47] Kyriaki Goulouti, Pierryves Padey, Alina Galimshina, Guillaume Habert, and Sébastien Lasvaux. Uncertainty of building elements' service lives in building LCA & LCC: What matters? *Building and Environment*, 183: 106904, 2020.

[48] Alex Ianchenko, Kathrina Simonen, and Clayton Barnes. Residential building lifespan and community turnover. *Journal of Architectural Engineering*, 26(3):04020026, 2020.

[49] Michele Ferreira Dias Morales, Natalia Reguly, Ana Paula Kirchheim, and Ana Passuello. Uncertainties related to the replacement stage in LCA of buildings: A case study of a structural masonry clay hollow brick wall. *Journal of Cleaner Production*, 251:119649, 2020.

[50] Hannes Harter, Manav Mahan Singh, Patricia Schneider-Marin, Werner Lang, and Philipp Geyer. Uncertainty analysis of life cycle energy assessment in early stages of design. *Energy and Buildings*, 208:109635, 2020.

[51] Shahaboddin Resalati, Christopher C Kendrick, and Callum Hill. Embodied energy data implications for optimal specification of building envelopes. *Building Research & Information*, 48(4):429–445, 2020.

[52] Mehdi Robati, Daniel Daly, and Georgios Kokogiannakis. A method of uncertainty analysis for whole-life embodied carbon emissions (CO_2-e) of building materials of a net-zero energy building in australia. *Journal of Cleaner Production*, 225:541–553, 2019.

[53] Peter Ylmén, Johanna Berlin, Kristina Mjörnell, and Jesper Arfvidsson. Managing choice uncertainties in life-cycle assessment as a decision-support tool for building design: A case study on building framework. *Sustainability*, 12(12):5130, 2020.

[54] Shu Su, Xiaodong Li, and Yimin Zhu. Dynamic assessment elements and their prospective solutions in dynamic life cycle assessment of buildings. *Building and Environment*, 158:248–259, 2019.

[55] Lucas Rosse Caldas, Anna Bernstad Saraiva, Vanessa Maria Andreola, and Romildo Dias Toledo Filho. Bamboo bio-concrete as an alternative for buildings' climate change mitigation and adaptation. *Construction and Building Materials*, 263:120652, 2020.

[56] Shu Su, Chen Zhu, and Xiaodong Li. A dynamic weighting system considering temporal variations using the dtt approach in lca of buildings. *Journal of Cleaner Production*, 220:398–407, 2019.

[57] Vladimir Zieger, Thibaut Lecompte, and Arthur Hellouin de Menibus. Impact of GHGs temporal dynamics on the GWP assessment of building materials: A case study on bio-based and non-bio-based walls. *Building and Environment*, 185:107210, 2020.

[58] Koji Negishi, Alexandra Lebert, Denise Almeida, Jacques Chevalier, and Ligia Tiruta-Barna. Evaluating climate change pathways through a building's lifecycle based on dynamic life cycle assessment. *Building and Environment*, 164:106377, 2019.

[59] Enrico Sicignano, Giacomo Di Ruocco, and Roberta Melella. Mitigation strategies for reduction of embodied energy and carbon, in the construction systems of contemporary quality architecture. *Sustainability*, 11(14):3806, 2019.

[60] F Asdrubali, I Ballarini, V Corrado, L Evangelisti, G Grazieschi, and C Guattari. Energy and environmental payback times for an NZEB retrofit. *Building and Environment*, 147:461–472, 2019.

[61] Carine Lausselet, Johana Paola Forero Urrego, Eirik Resch, and Helge Brattebø. Temporal analysis of the material flows and embodied greenhouse gas emissions of a neighborhood building stock. *Journal of Industrial Ecology*, 25(2):419–434, 2021.

[62] A. Kylili and P. A. Fokaides. Construction materials for the urban environment: Environmental assessment of life cycle performance. In Leticia Myriam Torres Martínez, Oxana Vasilievna Kharissova, and Boris Ildusovich Kharisov, editors, *Handbook of Ecomaterials*, pages 1–33. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-48281-1_133-1.

[63] Callum Aidan Stephen Hill. The environmental consequences concerning the use of timber in the built environment. *Frontiers in Built Environment*, 5:129, 2019.

[64] Diego Alexis Ramos Huarachi, Giovanna Goncalves, Antonio Carlos de Francisco, Maria Helene Giovanetti Canteri, and Cassiano Moro Piekarski. Life cycle assessment of traditional and alternative bricks: A review. *Environmental Impact Assessment Review*, 80:106335, 2020.

[65] Kate Krueger, Adam Stoker, and Gabrielle Gaustad. “alternative” materials in the green building and construction sector: Examples, barriers, and environmental analysis. *Smart and Sustainable Built Environment*, 2019.

[66] Rahul Dandautiya and Ajit Pratap Singh. Utilization potential of fly ash and copper tailings in concrete as partial replacement of cement along with life cycle assessment. *Waste Management*, 99:90–101, 2019.

[67] Ethan Ellingboe, Jay H Arehart, and Wil V Srubar. On the theoretical (CO_2) sequestration potential of pervious concrete. *Infrastructures*, 4(1): 12, 2019.

[68] Rafael Horn, Stefan Albrecht, Walter Haase, Max Langer, Daniel Schmeer, Werner Sobek, Olga Speck, and Philip Leistner. Bio-inspiration as a concept for sustainable constructions illustrated on graded concrete. *Journal of Bionic Engineering*, 16(4):742–753, 2019.

[69] Rawaz Kurda, Jorge de Brito, and José D Silvestre. Concretop method: Optimization of concrete with various incorporation ratios of fly ash and recycled aggregates in terms of quality performance and life-cycle cost and environmental impacts. *Journal of Cleaner Production*, 226:642–657, 2019.

[70] Amir Oladazimi, Saeed Mansour, and Seyed Abbas Hosseinijou. Comparative life cycle assessment of steel and concrete construction frames: A case study of two residential buildings in Iran. *Buildings*, 10(3):54, 2020.

[71] Yifei Shi, Yue Li, Yuzhou Tang, Xueliang Yuan, Qingsong Wang, Jinglan Hong, and Jian Zuo. Life cycle assessment of autoclaved aerated fly ash and concrete block production: A case study in China. *Environmental Science and Pollution Research*, 26(25):25432–25444, 2019.

[72] Vojtěch Václavík, Marcela Ondová, Tomáš Dvorskỳ, Adriana Eštoková, Martina Fabiánová, and Lukáš Gola. Sustainability potential evaluation of concrete with steel slag aggregates by the LCA method. *Sustainability*, 12 (23):9873, 2020.

[73] Huinan Wei, Ao Zhou, Tiejun Liu, Dujian Zou, and Hongshu Jian. Dynamic and environmental performance of eco-friendly ultra-high performance concrete containing waste cathode ray tube glass as a substitution of river sand. *Resources, Conservation and Recycling*, 162:105021, 2020.

[74] Sarah J Welsh-Huggins, Abbie B Liel, and Sherri M Cook. Reduce, reuse, resilient? life-cycle seismic and environmental performance of buildings with

alternative concretes. *Journal of Infrastructure Systems*, 26(1):04019033, 2020.

[75] Cindy X Chen, Francesca Pierobon, and Indroneil Ganguly. Life cycle assessment (LCA) of cross-laminated timber (CLT) produced in western washington: The role of logistics and wood species mix. *Sustainability*, 11 (5):1278, 2019.

[76] Katsuyuki Nakano, Masahiko Karube, and Nobuaki Hattori. Environmental impacts of building construction using cross-laminated timber panel construction method: A case of the research building in Kyushu, Japan. *Sustainability*, 12(6):2220, 2020.

[77] Katsuyuki Nakano, Wataru Koike, Ken Yamagishi, and Nobuaki Hattori. Environmental impacts of cross-laminated timber production in Japan. *Clean Technologies and Environmental Policy*, 22(10):2193–2205, 2020.

[78] Francesca Pierobon, Monica Huang, Kathrina Simonen, and Indroneil Ganguly. Environmental benefits of using hybrid CLT structure in midrise non-residential construction: An LCA based comparative case study in the us pacific northwest. *Journal of Building Engineering*, 26:100862, 2019.

[79] Maureen Puettmann, Arijit Sinha, and Indroneil Ganguly. Life cycle energy and environmental impacts of cross laminated timber made with coastal douglas-fir. *Journal of Green Building*, 14(4):17–33, 2019.

[80] Mehmet Kadri Akyüz. Determining economic and environmental impact of insulation by thermoeconomic and life cycle assessment analysis for different climate regions of turkey. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 43(7):829–851, 2021.

[81] Nima Amani and Ehsan Kiaee. Developing a two-criteria framework to rank thermal insulation materials in nearly zero energy buildings using multi-objective optimization approach. *Journal of Cleaner Production*, 276: 122592, 2020.

[82] Dario Bottino-Leone, Marco Larcher, Daniel Herrera-Avellanosa, Franziska Haas, and Alexandra Troi. Evaluation of natural-based internal insulation

systems in historic buildings through a holistic approach. *Energy*, 181:521–531, 2019.

[83] Yannay Casas-Ledón, Karen Daza Salgado, Juan Cea, Luis E Arteaga-Pérez, and Cecilia Fuentealba. Life cycle assessment of innovative insulation panels based on eucalyptus bark fibers. *Journal of Cleaner Production*, 249: 119356, 2020.

[84] Roberta Di Bari, Rafael Horn, Björn Nienborg, Felix Klinker, Esther Kieseritzky, and Felix Pawelz. The environmental potential of phase change materials in building applications: A multiple case investigation based on life cycle assessment and building simulation. *Energies*, 13(12):3045, 2020.

[85] José D Silvestre, André MP Castelo, José JBC Silva, Jorge MCL de Brito, and Manuel D Pinheiro. Energy retrofitting of a buildings' envelope: Assessment of the environmental, economic and energy (3E) performance of a cork-based thermal insulating rendering mortar. *Energies*, 13(1):143, 2020.

[86] Menghua Sun, William B Haskell, Tsan Sheng Ng, Alvin WL Ee, and Harn Wei Kua. Selection of building thermal insulation materials using robust optimization. *The International Journal of Life Cycle Assessment*, 25(3):443–455, 2020.

[87] Alba Torres-Rivas, Carlos Pozo, Mariana Palumbo, Anna Ewertowska, Laureano Jiménez, and Dieter Boer. Systematic combination of insulation biomaterials to enhance energy and environmental efficiency in buildings. *Construction and Building Materials*, 267:120973, 2021.

[88] Maja Wiprächtiger, Melanie Haupt, Niko Heeren, Eliane Waser, and Stefanie Hellweg. A framework for sustainable and circular system design: Development and application on thermal insulation materials. *Resources, Conservation and Recycling*, 154:104631, 2020.

[89] Claudia Fabiani, Anna Laura Pisello, Marco Barbanera, and Luisa F Cabeza. Palm oil-based bio-pcm for energy efficient building applications: Multipurpose thermal investigation and life cycle assessment. *Journal of Energy Storage*, 28:101129, 2020.

[90] Christina A Konstantinidou, Werner Lang, Agis M Papadopoulos, and Mattheos Santamouris. Life cycle and life cycle cost implications of integrated phase change materials in office buildings. *International Journal of Energy Research*, 43(1):150–166, 2019.

[91] Fabrice Motte, Gilles Notton, Chr Lamnatou, Christian Cristofari, and Daniel Chemisana. Numerical study of pcm integration impact on overall performances of a highly building-integrated solar collector. *Renewable Energy*, 137:10–19, 2019.

[92] Dimitra Papadaki, Spyros Foteinis, Vasileios Binas, Margarita N Assimakopoulos, Theocharis Tsoutsos, and George Kiriakidis. A life cycle assessment of PCM and VIP in warm mediterranean climates and their introduction as a strategy to promote energy savings and mitigate carbon emissions. *AIMS Materials Science*, 6(6):944–959, 2019.

[93] Hessam AzariJafari, Mohammad Javad Taheri Amiri, Ali Ashrafian, Haleh Rasekh, Maedeh Javaheri Barforooshi, and Javad Berenjian. Ternary blended cement: an eco-friendly alternative to improve resistivity of high-performance self-consolidating concrete against elevated temperature. *Journal of Cleaner Production*, 223:575–586, 2019.

[94] Simge Çankaya and Beyhan Pekey. A comparative life cycle assessment for sustainable cement production in Turkey. *Journal of Environmental Management*, 249:109362, 2019.

[95] R Gettu, A Patel, V Rathi, S Prakasan, AS Basavaraj, S Palaniappan, and S Maity. Influence of supplementary cementitious materials on the sustainability parameters of cements and concretes in the indian context. *Materials and Structures*, 52(1):1–11, 2019.

[96] Wu-Jian Long, Jie-Lin Tao, Can Lin, Yu-cun Gu, Liu Mei, Hua-Bo Duan, and Feng Xing. Rheology and buildability of sustainable cement-based composites containing micro-crystalline cellulose for 3d-printing. *Journal of Cleaner Production*, 239:118054, 2019.

[97] Ian Vázquez-Rowe, Kurt Ziegler-Rodriguez, Jara Laso, Isabel Quispe, Rubén Aldaco, and Ramzy Kahhat. Production of cement in Peru: Un-

derstanding carbon-related environmental impacts and their policy implications. *Resources, Conservation and Recycling*, 142:283–292, 2019.

[98] Robson Zulcao, Joao Luiz Calmon, Thais Ayres Rebello, and Darli Rodrigues Vieira. Life cycle assessment of the ornamental stone processing waste use in cement-based building materials. *Construction and Building Materials*, 257:119523, 2020.

[99] Jorge Fernandes, Marco Peixoto, Ricardo Mateus, and Helena Gervásio. Life cycle analysis of environmental impacts of earthen materials in the portuguese context: Rammed earth and compressed earth blocks. *Journal of Cleaner Production*, 241:118286, 2019.

[100] Alexandra H Meek, Mohamed Elchalakani, Christopher TS Beckett, and Timothy Grant. Alternative stabilised rammed earth materials incorporating recycled waste and industrial by-products: Life cycle assessment. *Construction and Building Materials*, 267:120997, 2021.

[101] Abhilash Holur Narayanaswamy, Peter Walker, BV Venkatarama Reddy, Andrew Heath, and Daniel Maskell. Mechanical and thermal properties, and comparative life-cycle impacts, of stabilised earth building products. *Construction and Building Materials*, 243:118096, 2020.

[102] Alberto Vilches, Antonio Garcia-Martinez, and Benito Sanchez-Montanes. Life cycle assessment (LCA) of building refurbishment: A literature review. *Energy and Buildings*, 135:286–301, 2017.

[103] Xabat Oregi, Patxi Hernandez, and Rufino Hernandez. Analysis of life-cycle boundaries for environmental and economic assessment of building energy refurbishment projects. *Energy and Buildings*, 136:12–25, 2017.

[104] Alina Galimshina, Maliki Moustapha, Alexander Hollberg, Pierryves Padey, Sébastien Lasvaux, Bruno Sudret, and Guillaume Habert. Statistical method to identify robust building renovation choices for environmental and economic performance. *Building and Environment*, 183:107143, 2020.

[105] Matheus Belucio, Carla Rodrigues, Carlos Henggeler Antunes, Fausto Freire, and Luis C Dias. Eco-efficiency in early design decisions: A multimethodology approach. *Journal of Cleaner Production*, 283:124630, 2021.

[106] Francesco Pittau, Gabriele Lumia, Niko Heeren, Giuliana Iannaccone, and Guillaume Habert. Retrofit as a carbon sink: The carbon storage potentials of the EU housing stock. *Journal of Cleaner Production*, 214:365–376, 2019.

[107] Agneta Ghose, Massimo Pizzol, Sarah J McLaren, Mathieu Vignes, and David Dowdell. Refurbishment of office buildings in New Zealand: identifying priorities for reducing environmental impacts. *The International Journal of Life Cycle Assessment*, 24(8):1480–1495, 2019.

[108] TM Gulotta, M Cellura, F Guarino, and S Longo. A bottom-up harmonized energy-environmental models for europe (BOHEEME): A case study on the thermal insulation of the EU-28 building stock. *Energy and Buildings*, 231: 110584, 2021.

[109] Xabat Oregi, Rufino Javier Hernández, and Patxi Hernandez. Environmental and economic prioritization of building energy refurbishment strategies with life-cycle approach. *Sustainability*, 12(9):3914, 2020.

[110] Lerwen Liu and Seeram Ramakrishna. *An introduction to circular economy.* Springer, Gateway East, Singapore, 1st ed. 2021. edition, 2021. ISBN 981-15-8510-5.

[111] Serena Giorgi, Monica Lavagna, and Andrea Campioli. Circular economy and regeneration of building stock: Policy improvements, stakeholder networking and life cycle tools. *Regeneration of the Built Environment from a Circular Economy Perspective*, pages 291–301, 2020.

[112] Atta Ajayebi, Peter Hopkinson, Kan Zhou, Dennis Lam, Han-Mei Chen, and Yong Wang. Spatiotemporal model to quantify stocks of building structural products for a prospective circular economy. *Resources, Conservation and Recycling*, 162:105026, 2020.

[113] Ingrid Bertin, Romain Mesnil, Jean-Marc Jaeger, Adélaïde Feraille, and Robert Le Roy. A BIM-based framework and databank for reusing load-bearing structural elements. *Sustainability*, 12(8):3147, 2020.

[114] Jan Brütting, Gennaro Senatore, Mattias Schevenels, and Corentin Fivet. Optimum design of frame structures from a stock of reclaimed elements. *Frontiers in Built Environment*, 6:57, 2020.

[115] Roberto Minunno, Timothy O'Grady, Gregory M Morrison, and Richard L Gruner. Exploring environmental benefits of reuse and recycle practices: A circular economy case study of a modular building. *Resources, Conservation and Recycling*, 160:104855, 2020.

[116] Jan Fořt and Robert Černý. Transition to circular economy in the construction industry: Environmental aspects of waste brick recycling scenarios. *Waste Management*, 118:510–520, 2020.

[117] Deniz Üçer Erduran, Soofia Tahira Elias-Ozkan, and Aleksey Ulybin. Assessing potential environmental impact and construction cost of reclaimed masonry walls. *The International Journal of Life Cycle Assessment*, 25(1): 1–16, 2020.

[118] Sirje Vares, Petr Hradil, Michael Sansom, and Viorel Ungureanu. Economic potential and environmental impacts of reused steel structures. *Structure and Infrastructure Engineering*, 16(4):750–761, 2020.

[119] JR Gouveia, E Silva, TM Mata, A Mendes, NS Caetano, and AA Martins. Life cycle assessment of a renewable energy generation system with a vanadium redox flow battery in a NZEB household. *Energy Reports*, 6:87–94, 2020.

[120] Daniel González-Prieto, Yolanda Fernández-Nava, Elena Marañón, and Maria Manuela Prieto. Influence of atlantic microclimates in northern spain on the environmental performance of lightweight concrete single-family houses. *Energies*, 13(17):4337, 2020.

[121] SSS Gardezi and N Shafiq. Operational carbon footprint prediction model for conventional tropical housing: A malaysian prospective. *International Journal of Environmental Science and Technology*, 16(12):7817–7826, 2019.

[122] Julien Walzberg, Thomas Dandres, Nicolas Merveille, Mohamed Cheriet, and Réjean Samson. Should we fear the rebound effect in smart homes? *Renewable and Sustainable Energy Reviews*, 125:109798, 2020.

[123] Eric O'Rear, David Webb, Joshua Kneifel, and Cheyney O'Fallon. Gas vs electric: Heating system fuel source implications on low-energy single-family

dwelling sustainability performance. *Journal of Building Engineering*, 25: 100779, 2019.

[124] Annick Anctil, Eunsang Lee, and Richard R Lunt. Net energy and cost benefit of transparent organic solar cells in building-integrated applications. *Applied Energy*, 261:114429, 2020.

[125] Gianluca Grazieschi, Paola Gori, Lidia Lombardi, and Francesco Asdrubali. Life cycle energy minimization of autonomous buildings. *Journal of Building Engineering*, 30:101229, 2020.

[126] Victor Kouloumpis, Antonios Kalogerakis, Anastasia Pavlidou, George Tsinarakis, and George Arampatzis. Should photovoltaics stay at home? Comparative life cycle environmental assessment on roof-mounted and ground-mounted photovoltaics. *Sustainability*, 12(21):9120, 2020.

[127] Hongyan Li, Yuehong Bi, Lifeng Qin, and Gaoli Zang. Absorption solar-ground source heat pump: Life cycle environmental profile and comparisons. *Geothermics*, 87:101850, 2020.

[128] José Adolfo Lozano Miralles, Rafael López García, José Manuel Palomar Carnicero, and Francisco Javier Rey Martínez. Comparative study of heat pump system and biomass boiler system to a tertiary building using the life cycle assessment (lca). *Renewable Energy*, 152:1439–1450, 2020.

[129] Simona Marinelli, Francesco Lolli, Maria Angela Butturi, Bianca Rimini, and Rita Gamberini. Environmental performance analysis of a dual-source heat pump system. *Energy and Buildings*, 223:110180, 2020.

[130] Georgios Martinopoulos. Are rooftop photovoltaic systems a sustainable solution for Europe? A life cycle impact assessment and cost analysis. *Applied Energy*, 257:114035, 2020.

[131] Mahdi Mehrtash, Florin Capitanescu, Per Kvols Heiselberg, Thomas Gibon, and Alexandre Bertrand. An enhanced optimal PV and battery sizing model for zero energy buildings considering environmental impacts. *IEEE Transactions on Industry Applications*, 56(6):6846–6856, 2020.

[132] Barbara Mendecka, Laura Tribioli, and Raffaello Cozzolino. Life cycle assessment of a stand-alone solar-based polygeneration power plant for a commercial building in different climate zones. *Renewable Energy*, 154:1132–1143, 2020.

[133] Zhuobiao Ni, Yue Wang, Yafei Wang, Shaoqing Chen, Manxi Xie, Tim Grotenhuis, and Rongliang Qiu. Comparative life-cycle assessment of aquifer thermal energy storage integrated with in situ bioremediation of chlorinated volatile organic compounds. *Environmental Science & Technology*, 54(5):3039–3049, 2020.

[134] Federico Rossi, Miguel Heleno, Riccardo Basosi, and Adalgisa Sinicropi. Environmental and economic optima of solar home systems design: A combined LCA and LCC approach. *Science of The Total Environment*, 744: 140569, 2020.

[135] Federico Rossi, Maria Laura Parisi, Sarah Greven, Riccardo Basosi, and Adalgisa Sinicropi. Life cycle assessment of classic and innovative batteries for solar home systems in Europe. *Energies*, 13(13):3454, 2020.

[136] NBIMS-US:2015(V3). NBIMS National BIM Standard - United States V3 - Terms and Definitions. Standard, 2015.

[137] Rafael Sacks, Chuck Eastman, Ghang Lee, and Paul Teicholz. *BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers.* John Wiley & Sons, 2018.

[138] Tiziano Dalla Mora, Erika Bolzonello, Carmine Cavalliere, and Fabio Peron. Key parameters featuring BIM-LCA integration in buildings: A practical review of the current trends. *Sustainability*, 12(17):7182, 2020.

[139] Alejandro Martínez-Rocamora, Carlos Rivera-Gómez, Carmen Galán-Marín, and Madelyn Marrero. Environmental benchmarking of building typologies through BIM-based combinatorial case studies. *Automation in Construction*, 132:103980, 2021.

[140] L Wastiels and R Decuypere. Identification and comparison of LCA-BIM integration strategies. In *IOP Conference Series: Earth and Environmental Science*, volume 323, page 012101. IOP Publishing, 2019.

[141] Raja Shahmir Nizam, Cheng Zhang, and Lu Tian. A BIM based tool for assessing embodied energy for buildings. *Energy and Buildings*, 170:1–14, 2018.

[142] Kaveh Safari and Hessam AzariJafari. Challenges and opportunities for integrating BIM and LCA: methodological choices and framework development. *Sustainable Cities and Society*, page 102728, 2021.

[143] Tajda Potrč Obrecht, Martin Röck, Endrit Hoxha, and Alexander Passer. BIM and LCA integration: A systematic literature review. *Sustainability*, 12(14):5534, 2020.

[144] Shu Su, Shimeng Li, Jingyi Ju, Qian Wang, and Zhao Xu. A building information modeling-based tool for estimating building demolition waste and evaluating its environmental impacts. *Waste Management*, 134:159–169, 2021.

[145] Saman Abbasi and Esmatullah Noorzai. The BIM-based multi-optimization approach in order to determine the trade-off between embodied and operation energy focused on renewable energy use. *Journal of Cleaner Production*, 281:125359, 2021.

[146] Patricia Schneider-Marin, Hannes Harter, Konstantin Tkachuk, and Werner Lang. Uncertainty analysis of embedded energy and greenhouse gas emissions using BIM in early design stages. *Sustainability*, 12(7):2633, 2020.

[147] Mona Abouhamad and Metwally Abu-Hamd. Life cycle assessment framework for embodied environmental impacts of building construction systems. *Sustainability*, 13(2):461, 2021.

[148] José Pedro Carvalho, Fernanda Schmitd Villaschi, and Luís Bragança. Assessing life cycle environmental and economic impacts of building construction solutions with BIM. *Sustainability*, 13(16):8914, 2021.

[149] S Theißen, J Höper, J Drzymalla, et al. Using open BIM and IFC to enable a comprehensive consideration of building services within a whole-building LCA. *Sustainability*, 12(14):5644, 2020.

[150] Husam Sameer and Stefan Bringezu. Building information modelling application of material, water, and climate footprint analysis. *Building Research & Information*, pages 1–20, 2021.

[151] Quddus Tushar, Muhammed A Bhuiyan, Guomin Zhang, and Tariq Maqsood. An integrated approach of BIM-enabled LCA and energy simulation: The optimized solution towards sustainable development. *Journal of Cleaner Production*, 289:125622, 2021.

[152] Cátia Raposo, Fernanda Rodrigues, and Hugo Rodrigues. BIM-based LCA assessment of seismic strengthening solutions for reinforced concrete precast industrial buildings. *Innovative Infrastructure Solutions*, 4(1):1–10, 2019.

[153] Mohammad Najjar, Karoline Figueiredo, Ahmed WA Hammad, and Assed Haddad. Integrated optimization with building information modeling and life cycle assessment for generating energy efficient buildings. *Applied Energy*, 250:1366–1382, 2019.

[154] Farzad Jalaei, Geoffrey Guest, Abhishek Gaur, and Jieying Zhang. Exploring the effects that a non-stationary climate and dynamic electricity grid mix has on whole building life cycle assessment: A multi-city comparison. *Sustainable Cities and Society*, 61:102294, 2020.

[155] Christina Kiamili, Alexander Hollberg, and Guillaume Habert. Detailed assessment of embodied carbon of hvac systems for a new office building based on BIM. *Sustainability*, 12(8):3372, 2020.

[156] Alexander Hollberg, Gianluca Genova, and Guillaume Habert. Evaluation of BIM-based LCA results for building design. *Automation in Construction*, 109:102972, 2020.

[157] Cristiane Bueno and Márcio Minto Fabricio. Comparative analysis between a complete LCA study and results from a BIM-LCA plug-in. *Automation in Construction*, 90:188–200, 2018.

[158] Sami G Al-Ghamdi and Melissa M Bilec. Green building rating systems and whole-building life cycle assessment: Comparative study of the existing assessment tools. *Journal of Architectural Engineering*, 23(1):04016015, 2017.

[159] K Forth, A Braun, and A Borrmann. BIM-integrated LCA-model analysis and implementation for practice. In *IOP Conference Series: Earth and Environmental Science*, volume 323, page 012100. IOP Publishing, 2019.

[160] Rosaliya Kurian, Kishor Sitaram Kulkarni, Prasanna Venkatesan Ramani, Chandan Swaroop Meena, Ashok Kumar, and Raffaello Cozzolino. Estimation of carbon footprint of residential building in warm humid climate of india through bim. *Energies*, 14(14):4237, 2021.

[161] Ruben Santos, António Aguiar Costa, José D Silvestre, and Lincy Pyl. Development of a BIM-based environmental and economic life cycle assessment tool. *Journal of Cleaner Production*, 265:121705, 2020.

[162] Rafael Horn, Sebastian Ebertshäuser, Roberta Di Bari, Olivia Jorgji, René Traunspurger, and Petra von Both. The BIM2LCA approach: An industry foundation classes (IFC)-based interface to integrate life cycle assessment in integral planning. *Sustainability*, 12(16):6558, 2020.

[163] Carmine Cavalliere, Guido Raffaele Dell'Osso, Alessandra Pierucci, and Francesco Iannone. Life cycle assessment data structure for building information modelling. *Journal of Cleaner Production*, 199:193–204, 2018.

[164] Cristiane Bueno, Lucas Melchiori Pereira, and Márcio Minto Fabricio. Life cycle assessment and environmental-based choices at the early design stages: An application using building information modelling. *Architectural Engineering and Design Management*, 14(5):332–346, 2018.

[165] Cristiane Bueno and Márcio Minto Fabricio. Methodological discussion of insertion and exportation of LCA data embedded in BIM elements. *WIT Transactions on the Built Environment*, 169:101–110, 2017.

[166] Mark Kyeredey Ansah, Xi Chen, Hongxing Yang, Lin Lu, and Patrick TI Lam. Developing an automated BIM-based life cycle assessment approach for modularly designed high-rise buildings. *Environmental Impact Assessment Review*, 90:106618, 2021.

[167] Sungwoo Lee, Sungho Tae, Hyungjae Jang, Chang U Chae, and Youngjin Bok. Development of building information modeling template for environmental impact assessment. *Sustainability*, 13(6):3092, 2021.

[168] Guiwen Liu, Tingyan Gu, Pengpeng Xu, Jingke Hong, Asheem Shrestha, and Igor Martek. A production line-based carbon emission assessment model for prefabricated components in china. *Journal of Cleaner Production*, 209:30–39, 2019.

[169] Qiang Du, Tana Bao, Yi Li, Youdan Huang, and Long Shao. Impact of prefabrication technology on the cradle-to-site CO_2 emissions of residential buildings. *Clean Technologies and Environmental Policy*, 21(7):1499–1514, 2019.

[170] Jian Li Hao, Baoquan Cheng, Weisheng Lu, Jun Xu, Junjie Wang, Weicheng Bu, and Zhiping Guo. Carbon emission reduction in prefabrication construction during materialization stage: A BIM-based life-cycle assessment approach. *Science of the Total Environment*, 723:137870, 2020.

[171] Yih Yoong Lip, Fang Yenn Teo, and Ioannes Yu Hoe Tang. Carbon footprint analysis of industrialised building system in Malaysia. In *AWAM International Conference on Civil Engineering*, pages 817–825. Springer, 2019.

[172] Fuyi Yao, Guiwen Liu, Yingbo Ji, Wenjing Tong, Xiaoyun Du, Kaijian Li, Asheem Shrestha, and Igor Martek. Evaluating the environmental impact of construction within the industrialized building process: a monetization and building information modelling approach. *International Journal of Environmental Research and Public Health*, 17(22):8396, 2020.

[173] Verena Göswein, Carla Rodrigues, José D Silvestre, Fausto Freire, Guillaume Habert, and Jakob König. Using anticipatory life cycle assessment to enable future sustainable construction. *Journal of Industrial Ecology*, 24 (1):178–192, 2020.

[174] Chunbo Zhang, Mingming Hu, Xining Yang, Arianna Amati, and Arnold Tukker. Life cycle greenhouse gas emission and cost analysis of prefabricated concrete building façade elements. *Journal of Industrial Ecology*, 24(5): 1016–1030, 2020.

[175] Ali Tighnavard Balasbaneh and Mohd Zamri Ramli. A comparative life cycle assessment (LCA) of concrete and steel-prefabricated prefinished volumetric construction structures in Malaysia. *Environmental Science and Pollution Research*, 27(34):43186–43201, 2020.

[176] Gholamreza Heravi, Milad Rostami, and Majid Fazeli Kebria. Energy consumption and carbon emissions assessment of integrated production and erection of buildings' pre-fabricated steel frames using lean techniques. *Journal of Cleaner Production*, 253:120045, 2020.

[177] Aisan Kong, Haibo Kang, Siyuan He, Na Li, and Wei Wang. Study on the carbon emissions in the whole construction process of prefabricated floor slab. *Applied Sciences*, 10(7):2326, 2020.

[178] Ester Pujadas-Gispert, David Sanjuan-Delmás, Albert de la Fuente, SPG Faas Moonen, and Alejandro Josa. Environmental analysis of concrete deep foundations: Influence of prefabrication, concrete strength, and design codes. *Journal of Cleaner Production*, 244:118751, 2020.

[179] Isolda Agustí-Juan, Andrei Jipa, and Guillaume Habert. Environmental assessment of multi-functional building elements constructed with digital fabrication techniques. *The International Journal of Life Cycle Assessment*, 24(6):1027–1039, 2019.

[180] Mohammad Kamali, Kasun Hewage, and Rehan Sadiq. Conventional versus modular construction methods: A comparative cradle-to-gate LCA for residential buildings. *Energy and Buildings*, 204:109479, 2019.

[181] Fernanda Belizario Silva, Daniel Costa Reis, Yazmin Lisbeth Mack-Vergara, Lucas Pessoto, Haibo Feng, Sérgio Almeida Pacca, Sébastien Lasvaux, Guillaume Habert, and Vanderley Moacyr John. Primary data priorities for the life cycle inventory of construction products: Focus on foreground processes. *The International Journal of Life Cycle Assessment*, 25:980–997, 2020.

[182] Robert H Crawford, André Stephan, and Fabian Prideaux. A comprehensive database of environmental flow coefficients for construction materials: Closing the loop in environmental design. In *Revisiting the Role of Architecture for 'Surviving' Development. 53rd International Conference of the Architectural Science Association*, pages 353–362. The Architectural Science Association, 2019.

[183] Mohammed Alzard, Hilal El-Hassan, and Tamer El Maaddawy. Development of a life cycle inventory dataset for recycled concrete aggregates in the city of Abu Dhabi. In *International Conference on Civil, Structural*

and Transportation Engineering, pages 227–228. International ASET Inc., 2020.

[184] Leslie Ayagapin and Jean Philippe Praene. Environmental overcost of single family houses in insular context: A comparative LCA study of Reunion Island and France. *Sustainability*, 12(21):8937, 2020.

[185] Victor Alberto Arvizu-Piña, Albert Cuchí-Burgos, and Itzia Gabriela Barrera-Alarcón. A top-down approach for implementation of environmental product declarations in Mexico's housing sector. *The International Journal of Life Cycle Assessment*, 25(1):157–167, 2020.

[186] Asger Alexander Wendt Karl, Esmir Maslesa, and Morten Birkved. Environmental performance assessment of the use stage of buildings using dynamic high-resolution energy consumption and data on grid composition. *Building and Environment*, 147:97–107, 2019.

[187] Matan Mayer and Martin Bechthold. Data granularity for life cycle modelling at an urban scale. *Architectural Science Review*, 63(3-4):351–360, 2020.

[188] Didier Vuarnoz, Endrit Hoxha, Julien Nembrini, Thomas Jusselme, and Stefano Cozza. Assessing the gap between a normative and a reality-based model of building LCA. *Journal of Building Engineering*, 31:101454, 2020.

[189] Pilar Mercader-Moyano, Paula M Esquivias, and Radu Muntean. Eco-efficient analysis of a refurbishment proposal for a social housing. *Sustainability*, 12(17):6725, 2020.

[190] Emmanuel Shittu, Valentina Stojceska, Petra Gratton, and Maria Kolokotroni. Environmental impact of cool roof paint: case-study of house retrofit in two hot islands. *Energy and Buildings*, 217:110007, 2020.

[191] Suzana Domjan, Ciril Arkar, and Sašo Medved. A computer-aided decision supporting tool for nearly zero energy building renovation. *Global Dwelling: Approaches to Sustainability, Design and Participation*, 193:177, 2020.

[192] Ayu Miyamoto, Karen Allacker, and Frank De Troyer. Visual tool to integrate LCA and LCC in the early design stage of housing. In *IOP Conference*

Series: Earth and Environmental Science, volume 323. IOP Publishing, 2019.

[193] Sandrine Duprez, Marine Fouquet, Quentin Herreros, and Thomas Jusselme. Improving life cycle-based exploration methods by coupling sensitivity analysis and metamodels. *Sustainable Cities and Society*, 44:70–84, 2019.

[194] TM Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, 1997.

[195] Ali Ghoroghi, Yacine Rezgui, Ioan Petri, and Thomas Beach. Advances in application of machine learning to life cycle assessment: A literature review. *The International Journal of Life Cycle Assessment*, 27(3):433–456, 2022.

[196] Natalia Nakamura Barros and Regina Coeli Ruschel. Machine learning for whole-building life cycle assessment: A systematic literature review. In *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering: ICCCBE 2020*, pages 109–122. Springer, 2021.

[197] Qian Shi and Yilong Xu. The selection of green building materials using GA-BP hybrid algorithm. In *2009 International Conference on Artificial Intelligence and Computational Intelligence*, volume 3, pages 40–45. IEEE, 2009.

[198] Antonio D'Amico, Giuseppina Ciulla, Marzia Traverso, V Lo Brano, and Elisabetta Palumbo. Artificial neural networks to assess energy and environmental performance of buildings: An Italian case study. *Journal of Cleaner Production*, 239:117993, 2019.

[199] Seyed Amirhosain Sharif Arani. *Optimizing energy performance of building renovation using traditional and machine learning approaches*. PhD thesis, Concordia University, 2020.

[200] Rahman Azari, Samira Garshasbi, Pegah Amini, Hazem Rashed-Ali, and Yousef Mohammadi. Multi-objective optimization of building envelope design for life cycle environmental performance. *Energy and Buildings*, 126: 524–534, 2016.

[201] Ping Hou, Olivier Jolliet, Ji Zhu, and Ming Xu. Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environment International*, 135:105393, 2020.

[202] Ping Hou. *Data-Driven Environmental System Analysis: Addressing Data Gaps in Life Cycle Assessment*. PhD thesis, University of Michigan, 2019.

[203] Andreas Froemelt, David J Durrenmatt, and Stefanie Hellweg. Using data mining to assess environmental impacts of household consumption behaviors. *Environmental Science & Technology*, 52(15):8467–8478, 2018.

[204] Mikaela Ann DeRousseau. *Concrete mixture design using machine learning, life cycle assessment, and multi-objective optimization*. PhD thesis, University of Colorado at Boulder, 2020.

[205] Xiangming Gu. *Metal-organic frameworks for post-combustion carbon capture-a life cycle assessment*. PhD thesis, The Ohio State University, 2018.

[206] Brandon Kuczenski, Christopher B Davis, Beatriz Rivela, and Krzysztof Janowicz. Semantic catalogs for life cycle assessment data. *Journal of Cleaner Production*, 137:1109–1117, 2016.

[207] Birte Glimm and Heiner Stuckenschmidt. 15 years of semantic web: An incomplete survey. *KI-Künstliche Intelligenz*, 30:117–130, 2016.

[208] Paul Anderson et al. *What is Web 2.0?: Ideas, technologies and implications for education*, volume 1. JISC Bristol, 2007.

[209] Ivan Herman. Semantic web adoption and applications, 2012. URL https://www.w3.org/People/Ivan/CorePresentations/Applications/. Accessed on January 8, 2023.

[210] John Davies, Rudi Studer, and Paul Warren. *Semantic Web technologies: trends and research in ontology-based systems*. John Wiley & Sons, 2006.

[211] Shelley Powers. *Practical RDF: solving problems with the resource description framework*. O'Reilly Media, Inc., 2003.

[212] Chris Welty, Deborah L McGuinness, and Michael K Smith. Owl web ontology language guide. *W3C recommendation, W3C (February 2004) http://www. w3. org/TR/2004/REC-owl-guide-20040210*, 48, 2004.

[213] F Henry Abanda, Joseph HM Tah, and Ramin Keivani. Trends in built environment semantic web applications: Where are we today? *Expert Systems with Applications*, 40(14):5563–5577, 2013.

[214] Birte Glimm, Chimezie Ogbuji, S Hawke, I Herman, B Parsia, A Polleres, and A Seaborne. SPARQL 1.1 entailment regimes. W3C recommendation 21 March 2013, 2013. URL https://www.w3.org/TR/sparql11-entailment/. Accessed on January 8, 2023.

[215] Jorge Pérez, Marcelo Arenas, and Claudio Gutierrez. Semantics and complexity of SPARQL. *ACM Transactions on Database Systems (TODS)*, 34 (3):1–45, 2009.

[216] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[217] Shaukat Ali and Shah Khusro. Poem: practical ontology engineering model for semantic web ontologies. *Cogent Engineering*, 3(1):1193959, 2016.

[218] Bo Liu, Keman Huang, Jianqiang Li, and MengChu Zhou. An incremental and distributed inference method for large-scale ontologies based on mapreduce paradigm. *IEEE Transactions on Cybernetics*, 45(1):53–64, 2014.

[219] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo. Methontology: from ontological art towards ontological engineering. In *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, pages 33–40. The Association for the Advancement of Artificial Intelligence (AAAI), 1997.

[220] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernandez-Lopez. The neon methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, 10(2):107–145, 2015.

[221] Krzysztof Janowicz, Adila Alfa Krisnadhi, Yingjie Hu, Sangwon Suh, Bo Pedersen Weidema, Beatriz Rivela, Johan Tivander, David E Meyer,

Gary Berg-Cross, Pascal Hitzler, et al. A minimal ontology pattern for life cycle assessment data. In *CEUR Workshop Proceedings*, volume 1461. CEUR-WS, 2015.

[222] Yingzhong Zhang, Xiaofang Luo, Jennifer J Buis, and John W Sutherland. Lca-oriented semantic representation for the product life cycle. *Journal of Cleaner Production*, 86:146–162, 2015.

[223] Bo Yan, Yingjie Hu, Brandon Kuczenski, Krzysztof Janowicz, Andrea Ballatore, Adila Alfa Krisnadhi, Yiting Ju, Pascal Hitzler, Sangwon Suh, and Wesley Ingwersen. An ontology for specifying spatiotemporal scopes in life cycle assessment. In *Diversity++@ ISWC*, pages 25–30, 2015.

[224] Benjamin Bertin, Vasile-Marian Scuturici, Emmanuel Risler, and Jean-Marie Pinon. A semantic approach to life cycle assessment applied on energy environmental impact data management. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 87–94, 2012.

[225] Benjamin Bertin, Vasile-Marian Scuturici, Jean-Marie Pinon, and Emmanuel Risler. Carbondb: a semantic life cycle inventory database. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 2683–2685, 2012.

[226] Vinit K Mittal, Sidney C Bailin, Michael A Gonzalez, David E Meyer, William M Barrett, and Raymond L Smith. Toward automated inventory modeling in life cycle assessment: The utility of semantic data modeling to predict real-world chemical production. *ACS Sustainable Chemistry & Engineering*, 6(2):1961–1976, 2018.

[227] Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE access*, 7, 2019.

[228] Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94, 2018.

[229] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital twin in manufacturing: A categorical literature review and classification. *Ifac-PapersOnline*, 51(11):1016–1022, 2018.

[230] Qinglin Qi, Fei Tao, Tianliang Hu, Nabil Anwer, Ang Liu, Yongli Wei, Lihui Wang, and AYC Nee. Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, 58:3–21, 2021.

[231] Martin Robert Enders and Nadja Hoßbach. Dimensions of digital twin applications: A literature review. In *Twenty-fifth Americas Conference on Information Systems*. Association for Information Systems, 2019.

[232] Calin Boje, Annie Guerriero, Sylvain Kubicki, and Yacine Rezgui. Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114, 2020.

[233] Fatemeh Nargesian, Erkang Zhu, Renée J Miller, Ken Q Pu, and Patricia C Arocena. Data lake management: challenges and opportunities. *Proceedings of the VLDB Endowment*, 12(12):1986–1989, 2019.

[234] Athira Nambiar and Divyansh Mundra. An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, 6(4):132, 2022.

[235] Matthias Jarke, Maurizio Lenzerini, Yannis Vassiliou, and Panos Vassiliadis. *Fundamentals of data warehouses*. Springer Science & Business Media, 2002.

[236] Bo Hu, Nuno Carvalho, Loredana Laera, and Takahide Matsutsuka. Towards big linked data: a large-scale, distributed semantic data storage. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*, pages 167–176, 2012.

[237] European Commission. A renovation wave for Europe: Greening our buildings, creating jobs, improving lives, 2020.

[238] Nina Campbell, L Ryan, V Rozite, E Lees, and G Heffner. Capturing the multiple benefits of energy efficiency. *IEA: Paris, France*, 2014.

[239] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.

[240] Hai-xiang Zhao and Frédéric Magoulès. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6): 3586–3592, 2012.

[241] The Intelligent Energy – Europe Programme of the European Union. *Implementing the Energy Performance of Buildings Directive (EPBD) – Featuring Country Reports*. European Union, 2015.

[242] Stefano Cozza, Jonathan Chambers, Chirag Deb, Jean-Louis Scartezzini, Arno Schlüter, and Martin K Patel. Do energy performance certificates allow reliable predictions of actual energy consumption and savings? Learning from the Swiss national database. *Energy and Buildings*, 224:110235, 2020.

[243] Ciara Ahern and Brian Norton. Energy performance certification: Misassessment due to assuming default heat losses. *Energy and Buildings*, 224: 110229, 2020.

[244] Y Li, S Kubicki, A Guerriero, and Y Rezgui. Review of building energy performance certification schemes towards future improvement. *Renewable and Sustainable Energy Reviews*, 113:109244, 2019.

[245] Stefano Cozza, Jonathan Chambers, Arianna Brambilla, and Martin K Patel. In search of optimal consumption: A review of causes and solutions to the energy performance gap in residential buildings. *Energy and Buildings*, 249:111253, 2021.

[246] Esfandiar Burman. *Assessing the operational performance of educational buildings against design expectations-a case study approach*. PhD thesis, UCL (University College London), 2016.

[247] Nishesh Jain, Esfand Burman, Samuel Stamp, Dejan Mumovic, and Michael Davies. Cross-sectoral assessment of the performance gap using calibrated building energy performance simulation. *Energy and Buildings*, 224:110271, 2020.

[248] Christine Eon, Jessica K Breadsell, Joshua Byrne, and Gregory M Morrison. The discrepancy between as-built and as-designed in energy efficient buildings: A rapid review. *Sustainability*, 12(16):6372, 2020.

[249] Patrick XW Zou, Dipika Wagle, and Morshed Alam. Strategies for minimizing building energy performance gaps between the design intend and the reality. *Energy and Buildings*, 191:31–41, 2019.

[250] Saleh Seyedzadeh, Farzad Pour Rahimian, Ivan Glesk, and Marc Roper. Machine learning for estimation of building energy consumption and performance: A review. *Visualization in Engineering*, 6(1):1–20, 2018.

[251] Muhammad Waseem Ahmad, Monjur Mourshed, and Yacine Rezgui. Trees vs neurons: Comparison between random forest and ann for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147:77–89, 2017.

[252] Xiwang Li and Jin Wen. Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, 37:517–537, 2014.

[253] Saeed Reza Mohandes, Xueqing Zhang, and Amir Mahdiyar. A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing*, 340:55–75, 2019.

[254] Muhammad Waseem Ahmad, Monjur Mourshed, Baris Yuce, and Yacine Rezgui. Computational intelligence techniques for HVAC systems: A review. *Building Simulation*, 9(4):359–398, 2016.

[255] Subodh Paudel, Mohamed Elmtiri, Wil L Kling, Olivier Le Corre, and Bruno Lacarrière. Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energy and Buildings*, 70: 81–93, 2014.

[256] Ioan Petri, Sylvain Kubicki, Yacine Rezgui, Annie Guerriero, and Haijiang Li. Optimizing energy efficiency in operating built environment assets through building information modeling: A case study. *Energies*, 10(8):1167, 2017.

[257] Chirag Deb, Lee Siew Eang, Junjing Yang, and Mattheos Santamouris. Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. *Energy and Buildings*, 121:284–297, 2016.

[258] Qiong Li, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10):2249–2256, 2009.

[259] Chao Ding and Khee Poh Lam. Data-driven model for cross ventilation potential in high-density cities based on coupled CFD simulation and machine learning. *Building and Environment*, 165:106394, 2019.

[260] Babak Raji, Martin J Tenpierik, Regina Bokel, and Andy van den Dobbelsteen. Natural summer ventilation strategies for energy-saving in high-rise buildings: A case study in the netherlands. *International Journal of Ventilation*, 19(1):25–48, 2020.

[261] Zheming Tong, Yujiao Chen, Ali Malkawi, Zhu Liu, and Richard B Freeman. Energy saving potential of natural ventilation in china: The impact of ambient air pollution. *Applied Energy*, 179:660–668, 2016.

[262] Elena Barbadilla-Martín, José Guadix Martín, José Manuel Salmerón Lissén, José Sánchez Ramos, and Servando Álvarez Domínguez. Assessment of thermal comfort and energy savings in a field study on adaptive comfort with application for mixed mode offices. *Energy and Buildings*, 167: 281–289, 2018.

[263] Maite Gil-Baez, Ángela Barrios-Padura, Marta Molina-Huelva, and R Chacartegui. Natural ventilation systems in 21st-century for near zero energy school buildings. *Energy*, 137:1186–1200, 2017.

[264] Yujiao Chen, Leslie K Norford, Holly W Samuelson, and Ali Malkawi. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings*, 169:195–205, 2018.

[265] Hansaem Park and Dong Yoon Park. Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms. *Building and Environment*, 195:107744, 2021.

[266] Kyosuke Hiyama, Kenichiro Takeuchi, Yuichi Omodaka, and Thanyalak Srisamranrungruang. Operation strategy for engineered natural ventilation using machine learning under sparse data conditions. *Japan Architectural Review*, 5(1):119–126, 2022.

[267] Ioanna Vrachimi, Ana Paula Melo, and Daniel Cóstola. Prediction of wind pressure coefficients in building energy simulation using machine learning. In *Proceedings of the 15th IBPSA Conference*, pages 2334–2341. International Building Performance Simulation Association, 2017.

[268] Wael A Yousef Mousa, Werner Lang, Thomas Auer, and Waleed A Yousef. A pattern recognition approach for modeling the air change rates in naturally ventilated buildings from limited steady-state CFD simulations. *Energy and Buildings*, 155:54–65, 2017.

[269] Yujiao Chen, Zheming Tong, Yang Zheng, Holly Samuelson, and Leslie Norford. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. *Journal of Cleaner Production*, 254:119866, 2020.

[270] Vincent JL Gan, Boyu Wang, CM Chan, AU Weerasuriya, and Jack CP Cheng. Physics-based, data-driven approach for predicting natural ventilation of residential high-rise buildings. *Building Simulation: An International Journal*, 15(1):129–148, 2022.

[271] Antonio J Aguilar, L María, Nélson Costa, Pedro Arezes, María D Martínez-Aires, and Diego P Ruiz. Assessment of ventilation rates inside educational buildings in southwestern Europe: Analysis of implemented strategic measures. *Journal of Building Engineering*, 51:104204, 2022.

[272] Tian-Wen Wang, Wei Yin, Lin-Li Fu, and Zhi-Yi Zhang. Estimation model for natural ventilation by wind force considering wind direction and building orientation for low-rise building in China. *Indoor and Built Environment*, page 1420326X20944983, 2020.

[273] Roberto Z Freire, Marc O Abadie, and Nathan Mendes. On the improvement of natural ventilation models. *Energy and Buildings*, 62:222–229, 2013.

[274] K Sheshagiri Hebbar, PA Paranjpe, and K Sridhara. Performance of conical jet nozzles in terms of discharge coefficient. *Aeronautical Society of India*, 22(1):3–9, 1970.

[275] Tsang-Chu Yu and Chung-Chih Lin. An intelligent wireless sensing and control system to improve indoor air quality: Monitoring, prediction, and preaction. *International Journal of Distributed Sensor Networks*, 11(8): 140978, 2015.

[276] B Khazaei, A Shiehbeigi, and AR Haji Molla Ali Kani. Modeling indoor air carbon dioxide concentration using artificial neural network. *International Journal of Environmental Science and Technology*, 16:729–736, 2019.

[277] Johanna Kallio, Jaakko Tervonen, Pauli Räsänen, Riku Mäkynen, Jani Koivusaari, and Johannes Peltola. Forecasting office indoor CO_2 concentration using machine learning with a one-year dataset. *Building and Environment*, 187:107409, 2021.

[278] Guangfei Yang, Erbiao Yuan, and Wenjun Wu. Predicting the long-term CO_2 concentration in classrooms based on the BO–EMD–LSTM model. *Building and Environment*, 224:109568, 2022.

[279] Marc Lotteau, Philippe Loubet, Maxime Pousse, Emmanuel Dufrasnes, and Guido Sonnemann. Critical review of life cycle assessment (LCA) for the built environment at the neighborhood scale. *Building and Environment*, 93:165–178, 2015.

[280] Alessio Mastrucci, Antonino Marvuglia, Ulrich Leopold, and Enrico Benetto. Life cycle assessment of building stocks from urban to transnational scales: A review. *Renewable and Sustainable Energy Reviews*, 74: 316–332, 2017.

[281] Magnus Österbring, Érika Mata, Liane Thuvander, and Holger Wallbaum. Explorative life-cycle assessment of renovating existing urban housing-stocks. *Building and Environment*, 165:106391, 2019.

[282] Karen Allacker, Valentina Castellani, Giorgio Baldinelli, Francesco Bianchi, Catia Baldassarri, and Serenella Sala. Energy simulation and LCA for

macro-scale analysis of eco-innovations in the housing stock. *The International Journal of Life Cycle Assessment*, 24(6):989–1008, 2019.

[283] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43 (5-6):907–928, 1995.

[284] Shaun Howell, Yacine Rezgui, Jean-Laurent Hippolyte, Bejay Jayan, and Haijiang Li. Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources. *Renewable and Sustainable Energy Reviews*, 77:193–214, 2017.

[285] Mark Saunders, Philip Lewis, and Adrian Thornhill. Research methods. *Business Students 6th edition Pearson Education Limited, England*, 2012.

[286] World Health Organization. *Health Research Methodology: A Guide for Training in Research Methods Second Edition*, volume 5. World Health Organization, 2001.

[287] Aleksandras Melnikovas. Towards an explicit research methodology: Adapting research onion model for futures studies. *Journal of Futures Studies*, 23 (2):29–44, 2018.

[288] Egon G Guba, Yvonna S Lincoln, et al. *Fourth generation evaluation*. Sage, 1989.

[289] Jane Ritchie, Jane Lewis, Carol McNaughton Nicholls, Rachel Ormston, et al. *Qualitative research practice: A guide for social science students and researchers*. sage, 2013.

[290] Marcia Mkansi and Edwin Asiamah Acheampong. Research philosophy debates and classifications: students' dilemma. *Electronic Journal of Business Research Methods*, 10(2):pp132–140, 2012.

[291] Mihaela L Kelemen and Nick Rumens. *An introduction to critical management research*. Sage, 2008.

[292] David A Kolb. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[293] Marian Petre. Uml in practice. In *2013 35th international conference on software engineering (icse)*, pages 722–731. IEEE, 2013.

[294] Grady Booch, James Rumbaugh, and Ivar Jacobson. *The Unified Modeling Language user guide*. Pearson Education, 2005.

[295] Nermeen Elkashef, Yasser F Hassan, et al. Mapping uml sequence diagram into the web ontology language owl. *International Journal of Advanced Computer Science and Applications*, 11(5), 2020.

[296] Michael D Myers and David Avison. *Qualitative research in information systems: a reader*. Sage, 2002.

[297] ASHRAE 62.1 - The Standards for Ventilation and Indoor Air Quality . Standard, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2022.

[298] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.

[299] Michel Verleysen and Damien François. The curse of dimensionality in data mining and time series prediction. In *International work-conference on artificial neural networks*, pages 758–770. Springer, 2005.

[300] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[301] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[302] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2): 197–227, 2016.

[303] Mohamed Abuella and Badrul Chowdhury. Random forest ensemble of support vector regression models for solar power forecasting. In *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2017.

[304] Ali Lahouar and J Ben Hadj Slama. Hour-ahead wind power forecast based on random forests. *Renewable Energy*, 109:529–541, 2017.

[305] Francesco Smarra, Achin Jain, Tullio De Rubeis, Dario Ambrosini, Alessandro D'Innocenzo, and Rahul Mangharam. Data-driven model predictive control using random forests for building energy optimization and climate control. *Applied Energy*, 226:1252–1272, 2018.

[306] Gongbo Chen, Yichao Wang, Shanshan Li, Wei Cao, Hongyan Ren, Luke D Knibbs, Michael J Abramson, and Yuming Guo. Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach. *Environmental Pollution*, 242:605–613, 2018.

[307] Foster Provost and Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc., 2013.

[308] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[309] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.

[310] Bohdan Macukow. Neural networks–state of art, brief history, basic models and architecture. In *IFIP international conference on computer information systems and industrial management*, pages 3–14. Springer, 2016.

[311] Ying Sun, Fariborz Haghighat, and Benjamin CM Fung. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings*, 221, 2020.

[312] John H Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[313] Tiejun Li, Guifang Shao, Wangda Zuo, and Sen Huang. Genetic algorithm for building optimization: State-of-the-art survey. In *Proceedings of the 9th international conference on machine learning and computing*, pages 205–210, 2017.

[314] Ines Costa-Carrapico, Rokia Raslan, and Javier Neila González. A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency. *Energy and Buildings*, 210:109690, 2020.

[315] ISO 14040:2006(EN). ISO 14040:2006(en) Environmental management — Life cycle assessment — Principles and framework. Standard, International Organization for Standardization, 2006.

[316] ISO 14044:2006(EN). ISO 14044:2006(en) Environmental management — Life cycle assessment — Requirements and guidelines. Standard, International Organization for Standardization, 2006.

[317] EN 15978:2011(EN). EN 15978:2011(en) Sustainability of construction works. Assessment of environmental performance of buildings. Calculation method. Standard, European Standards, 2011.

[318] EN 15804:2019(en). EN 15804:2019(en) Sustainability of construction works - Environmental product declarations - Core rules for the product category of construction products. Standard, European Standards, 2019.

[319] Environmental design - CIBSE Guide A . Standard, The Chartered Institution of Building Services Engineers, London, UK, 2021.

[320] Tracy A Jenkin, Lindsay McShane, and Jane Webster. Green information technologies and systems: Employees' perceptions of organizational practices. *Business & Society*, 50(2):266–314, 2011.

[321] Joseph Sarkis and Hanmin Zhu. Information technology and systems in china's circular economy: Implications for sustainability. *Journal of Systems and Information Technology*, 10(3):202–217, 2008.

[322] Anubha Jain, Manoj Mishra, Sateesh Kumar Peddoju, and Nitin Jain. Energy efficient computing-green cloud computing. In *2013 international conference on energy efficient technologies for sustainability*, pages 978–982. IEEE, 2013.

[323] Joseph Sarkis, Chulmo Koo, and Richard T Watson. Green information systems & technologies–this generation and beyond: Introduction to the special issue. *Information Systems Frontiers*, 15:695–704, 2013.

[324] Joyce Thomson and Tim Jackson. Sustainable procurement in practice: Lessons from local government. *Journal of Environmental Planning and Management*, 50(3):421–444, 2007.

[325] Nisha Rani Misra, Sandeep Kumar, and Arpit Jain. A review on e-waste: Fostering the need for green electronics. In *2021 international conference on computing, communication, and intelligent systems (ICCCIS)*, pages 1032–1036. IEEE, 2021.

[326] Samuel Ackerman, Orna Raz, Marcel Zalmanovici, and Aviad Zlotnick. Automatically detecting data drift in machine learning classifiers. *arXiv preprint arXiv:2111.05672*, 2021.

[327] Ankur Mallick, Kevin Hsieh, Behnaz Arzani, and Gauri Joshi. Matchmaker: Data drift mitigation in machine learning for large-scale systems. *Proceedings of Machine Learning and Systems*, 4:77–94, 2022.

[328] Eva García-Martín, Crefeda Faviola Rodrigues, Graham Riley, and Håkan Grahn. Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, 134:75–88, 2019.

[329] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. Green AI. *Communications of the ACM*, 63(12):54–63, 2020.

[330] Murad Qasaimeh, Kristof Denolf, Jack Lo, Kees Vissers, Joseph Zambreno, and Phillip H Jones. Comparing energy efficiency of cpu, gpu and fpga implementations for vision kernels. In *2019 IEEE international conference on embedded software and systems (ICESS)*, pages 1–8. IEEE, 2019.

[331] Michael A Laurenzano, Ananta Tiwari, Adam Jundt, Joshua Peraza, William A Ward, Roy Campbell, and Laura Carrington. Characterizing the performance-energy tradeoff of small ARM cores in HPC computation. In *Euro-Par 2014 Parallel Processing: 20th International Conference*, pages 124–137. Springer, 2014.

[332] Jérémy Bonvoisin, Alan Lelah, Fabrice Mathieux, and Daniel Brissaud. An environmental assessment method for wireless sensor networks. *Journal of Cleaner Production*, 33:145–154, 2012.

[333] Mohammad Mustafa Taye. Understanding semantic web and ontologies: Theory and applications. *arXiv preprint arXiv:1006.4567*, 2010.

[334] John Davies. Lightweight ontologies. In Roberto Poli, Michael Healy, and Achilles Kameas, editors, *Theory and Applications of Ontology: Computer Applications*, pages 197–229. Springer, Netherlands, 2010.

[335] Mike Uschold. Building ontologies: Towards a unified methodology. In *Proceedings of 16th Annual Conference of the British Computer Society Specialists Group on Expert Systems*. Citeseer, 1996.

[336] Michael Gruninger and Mark Fox. Methodology for the design and evaluation of ontologies. In *Proc. IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.

[337] York Sure, Steffen Staab, and Rudi Studer. On-to-knowledge methodology (otkm). *Handbook on ontologies*, pages 117–132, 2004.

[338] Antonio De Nicola, Michele Missikoff, and Roberto Navigli. A software engineering approach to ontology building. *Information Systems*, 34(2): 258–275, 2009.

[339] Mariano Fernández-López. Overview of methodologies for building ontologies. 1999.

[340] Steffen Staab, Rudi Studer, H-P Schnurr, and York Sure. Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34, 2001.

[341] Michael Compton, Payam Barnaghi, Luis Bermudez, Raul Garcia-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, et al. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.

[342] Simon JD Cox. Iso 19156: 2011-geographic information: Observations and measurements. *International Organization for Standardization*, 2011.

[343] Ralph Hodgson and Paul J Keller. Qudt-quantities, units, dimensions and data types in owl and xml. *Online (September 2011) http://www. qudt. org*, 34, 2011.

[344] Pieter Pauwels, Sijie Zhang, and Yong-Cheol Lee. Semantic web technologies in aec industry: A literature overview. *Automation in Construction*, 73:145–165, 2017.

Appendix

Appendix A

Use Cases Taxonomy

Through the literature review conducted in Chapter 2, and several workshops with industry experts, the use cases were found to have three major dimensions—that is, physical, temporal, and enabler—(as shown in Figure 1), which will be described in more detail below.

- The physical dimension: identifies the type of the asset to be evaluated and the scale of LCA application. There are four major categories: Building, Utility, Transport, and Open spaces, each of which is the root category for a set of sub-types of assets. A use case can either be applied to an individual asset or to multiple assets.
- The temporal dimension: includes all of the major life-cycle stages in which a specific use case can be applied. Three major sub-categories were identified, as follows: i) new build (e.g., assessment of design proposals of a new asset, an extension to an existing asset, or a comparison of major construction processes and construction systems), ii) building operation (e.g., energy consumption, and trade-offs between ventilation requirements and energy use), and iii) retrofit and renovation (e.g., structural repairs, and energy retrofit measures).
- The enabler: represents the required hardware and software to deliver the intended use case. This dimension is intended to identify the numerous tools and models that serve a specific goal (e.g., environmental assessment tools for conducting LCA calculations). It also uses IoT devices to collect data and controllers to actuate the system.

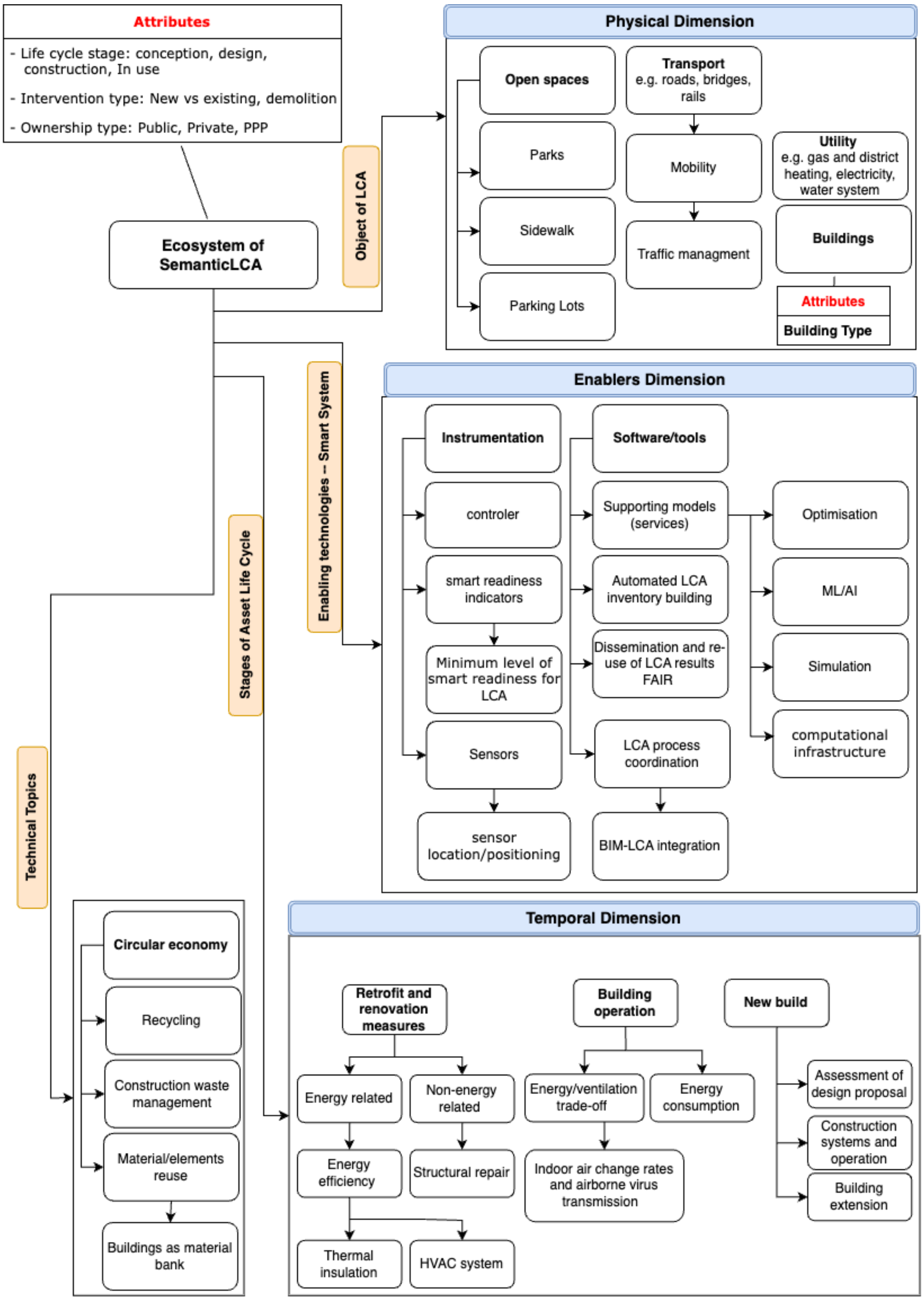


Figure 1: Use cases taxonomy

Appendix B

Semantisation of LCA Use Cases

A semantisation of use cases technique was developed, through which a use case can be analysed from eight different layers (as illustrated in Figure 2).

- Spatial scope, which aims to specify the physical boundaries of the asset(s) associated with the use case. The decision at this level will have several implications on the subsequent layers because the requirements and implementation approaches significantly vary.

- Domain, which identifies the type of the built asset (i.e., buildings, utility, transport, and open spaces).

- Life-cycle stage, which identifies the life-cycle stage that is associated with the use case. This aims to recognise the development stage of the asset (i.e., design, construction, in-use, and demolition).

- Scope of LCA, which relates to the scope and system boundary of the LCA study. It is possible for an LCA use case to include the entire product life-cycle stages, or it may be limited to particular stages. This layer should not be conflated with the previous layer. To clarify, a use case can only be associated with one project's life-cycle stage; for example, assessing design alternatives during the design stage, hence, specifying when the use case was initiated and executed. Meanwhile, a study must declare the product stages that should be considered as part of the LCA methodology. In the given example, during the evaluation of a design proposal, it is possible to include the manufacturing, construction, maintenance, and EoL treatment of the building materials associated with the proposed design.

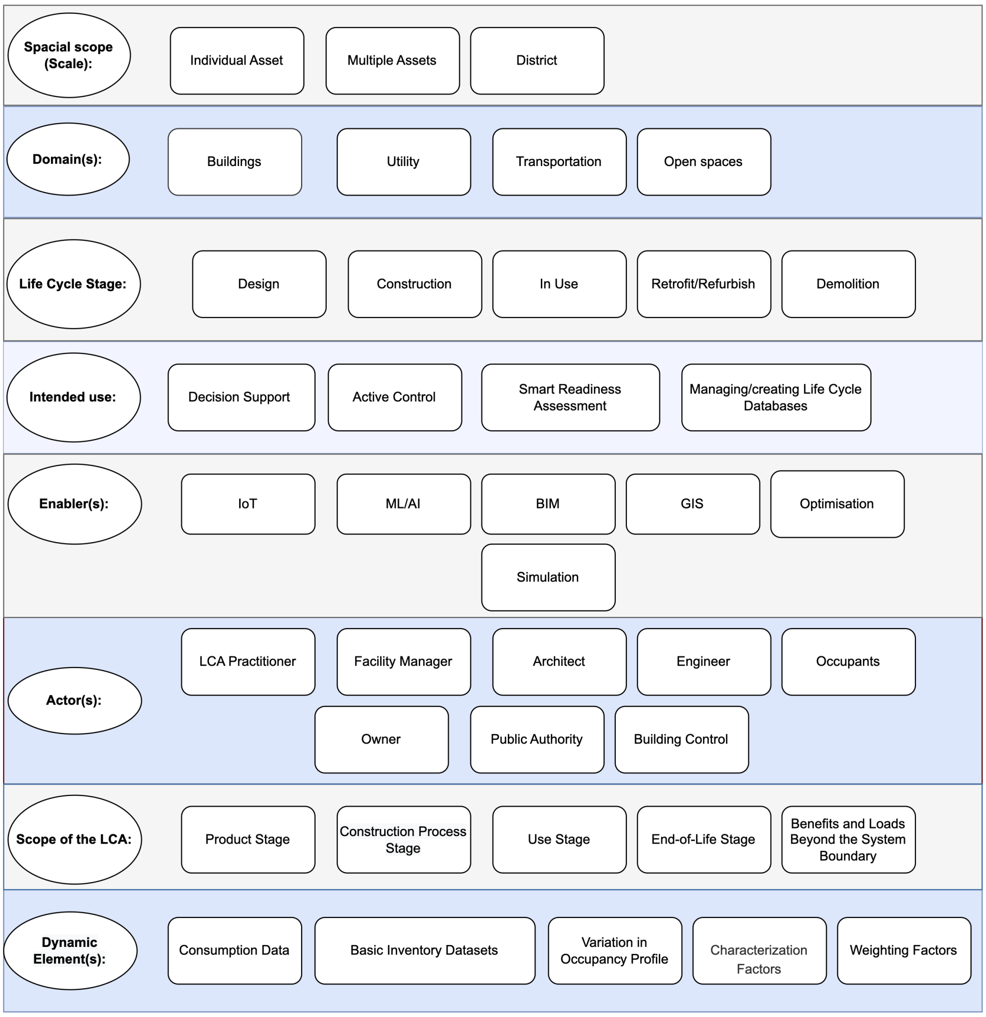


Figure 2: Semantisation of LCA use cases

- Intended use, which relates to the objective and intended applications of the use case. These objectives can run for longer periods of time (e.g., active control of building environmental performance) or for a certain period (e.g., the assessment of a design proposal).

- Enabler, which refers to the various digital resources that are used for the implementation of the use case. These include data collection sources, digital models (e.g., BIM), and modelling techniques (e.g., ML, optimisation, and simulation).

- Actor, which refers to the key players that are involved in either the development or execution of the use case. Players are assigned to different activities depending on their knowledge domain and responsibility.

- Dynamic element, which references the time-dependant factors when applying LCA. A use case can have a single or multiple dynamic elements, depending on the chosen scope and system boundary.

It is important to note here that the list of entities in each layer is not exhaustive, and therefore there is room for defining further entities.