Article | Received 29 June 2026; Revised 10 September 2026; Accepted 22 September 2026; Published 29 September 2026
https://doi.org/10.55092/sc20260019

Industry Foundation Classes (IFC)-enabled whole-building semantic optimisation and query-driven decision-support workflow for low-carbon and circular smart construction

Yiping Meng[1,*], Yiming Sun[2], Haoyu Huang[3] and Cong Zhang[4,*]

[1] School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, UK
[2] School of Electrical and Electronic Engineering, The University of Sheffield, Sheffield, UK
[3] School of Engineering, Newcastle University, Newcastle, UK
[4] College of Architecture and Civil Engineering, Beijing University of Technology, Beijing, China

* Correspondence authors; E-mails: y.meng@lboro.ac.uk (Y.M.); zhangcong@bjut.edu.cn (C.Z.).

Highlights:

- Open Industry Foundation Classes (IFC) files are parsed directly into optimisation-ready component inventories, with no dependence on a proprietary Application Programming Interface (API).
- Walls, floors, and roofs are enumerated together, so the Pareto ranking of candidate configurations picks up how the material chosen for one component interacts with the others.
- A set of 12 controlled, deterministic query types turns the numeric results into component-aware summaries of each candidate scenario.

Abstract: Material choices fixed early in design determine much of a building's embodied carbon and how much of its material can later be recovered, but the data needed to weigh the two are usually held in separate tools. To bring them together, a deterministic process was implemented that reads Industry Foundation Classes (IFC) data, enumerates whole-building material scenarios, filters them for Pareto dominance, answers controlled semantic queries, and packages the results as prompt text for optional downstream interpretation. One buildingSMART reference IFC model yielded seven valid components, and 150 wall-floor-roof material configurations were evaluated on that basis. Twelve configurations were non-dominated: six low-carbon-oriented configurations, in which timber systems predominate, and six circularity-oriented configurations built on aluminium-roof systems. The same candidates were returned under all five Building Circularity Score (BCS) weighting schemes (12/12 overlap with the equal-weight baseline), and the classification of 0 balanced, 6 low-carbon-oriented, and 6 circularity-oriented scenarios held across nine threshold combinations. In a scenario-based Monte Carlo check, every one of the 12 baseline candidates stayed non-dominated in at least 83.2% of 1000 sampled runs under the stated assumptions. The full 150-configuration case took approximately 0.79 s, of which pairwise Pareto filtering accounted for approximately 0.023 s. Within this single test case, IFC-derived quantities were sufficient for a reproducible environmental ranking, and the outputs are candidate scenarios for

prototype-stage decision support, not complete design prescriptions. External validation, additional IFC models, engineering feasibility checks, and deployment beyond the prototype are left to future work.

Keywords: Building Information Modelling (BIM); Industry Foundation Classes (IFC); whole-building optimisation; embodied carbon; circular construction; semantic query; multi-objective optimisation; Pareto analysis; deterministic prompt packaging

1. Introduction

1.1. Context and motivation

Much of the embodied carbon a building carries over its life is determined by material decisions taken early in design, yet carbon and circularity data are seldom at hand together when components are being chosen [1]. The stakes are considerable: the Global Alliance for Buildings and Construction (GlobalABC) attributes around 37% of global CO_2 emissions, 28% of global energy consumption, and nearly 50% of global material extraction to buildings and construction [2]. Regulation has historically concentrated on operational carbon. As envelopes become more energy-efficient, the emissions arising from material extraction, manufacturing, transport, construction, and end-of-life make up an increasingly important, and still unmitigated, share of whole-building life cycle impacts [3].

Building-scale assessment of environmental performance, and of sustainability more broadly, is framed by EN 15978 and ISO 21931-1 [4,5]. The scope of this study is deliberately narrower. The prototype reported here estimates cradle-to-gate embodied carbon for the purpose of comparing scenarios and pairs that estimate with a circularity proxy. It is not a full EN 15978-compliant whole-life carbon assessment, and it is not a complete ISO 21931 sustainability assessment.

It follows that material intervention has most leverage early in design. In their review of strategies for structural embodied carbon, Fang *et al.* [6] found both holistic and material-specific measures to be accessible and high-impact. The 109 case studies examined by Pomponi and Moncaster [7] point the same way: substituting materials late in design saves little compared with decisions made earlier at system level. Carbon and circularity therefore have to enter the process before material specifications are fixed.

Circular-economy principles, which aim to eliminate waste, keep products and materials in circulation, and regenerate natural systems, complement carbon reduction [8]. Construction is a priority sector for them, because buildings are long-lived assemblies whose components can in principle be designed for disassembly, reuse, and recycling [9]. Measuring circularity at building scale is harder. At product level, the Material Circularity Indicator (MCI) tracks recycled flows and retained material value, but on its own it does not capture building-specific concerns such as disassembly potential or information completeness [10]. Assessment has been extended to buildings through the Building Circularity Indicator (BCI), the Madaster Circularity Indicator (MAD-CI), and the Circular Building Assessment Prototype, although these remain detached from design-phase optimisation workflows [11].

The problem addressed here arises where these strands meet. Life cycle assessment (LCA) applications serve mainly for environmental accounting; the semantic information held in Building Information Modelling (BIM) models is rarely drawn on to compare materials; and circularity indicators tend to be applied outside the design workflow. Emerging large language model (LLM) applications offer a further interface, but they have no inherent connection to structured building data. The approach examined in this paper joins Industry Foundation Classes (IFC)-based extraction, multi-objective enumeration, and controlled queries so that these information sources can be used within one traceable process.

1.2. BIM, IFC, and semantic orchestration

Alongside component geometry, a BIM model carries material, quantity, and spatial information. IFC, which buildingSMART International maintains and which is standardised as ISO 16739, allows these data to be exchanged in an open, vendor-neutral form [12]. Building elements appear in the schema as entities such as IfcWall, IfcSlab, IfcRoof, and IfcColumn. Spatial containment is expressed through IfcBuildingStorey, quantities through IfcElementQuantity, and materials through IfcMaterial and IfcMaterialLayerSet. A component-level carbon and circularity calculation depends on exactly these records [13].

Starting from IFC means that the calculation begins with the model as exchanged, not with a particular authoring tool. Material systems, quantities, and spatial relationships can then be extracted and mapped without a vendor-specific application programming interface (API) [14]. Proprietary APIs do integrate closely with their platforms, but they bring licensing, runtime, and specialist-knowledge requirements. Our implementation reads standardised STEP (Standard for the Exchange of Product model data) Physical File (SPF) records directly, which keeps processing deterministic and removes the need for a proprietary runtime. bimSPARQL takes a related route, using Resource Description Framework (RDF) representations of BIM data to support structured queries [15].

IFC is used here as one layer among the openBIM information-management standards, not as a replacement for them. Checkable IFC information requirements are expressed through the Information Delivery Specification (IDS); shared terms for classifications, properties, and materials come from the buildingSMART Data Dictionary (bSDD); and Level of Information Need (LOIN) standards set out what information is required at an exchange point [16–18]. The schema versions IFC 4.3, IFC4, and IFC2x3 also differ in their implementation histories [12]. Our prototype reads only the selected buildingSMART reference model. It carries out no IDS checks, bSDD material classification, formal LOIN assessment, or benchmarking across IFC versions; these are possible alignment tasks, not capabilities claimed for the present implementation.

Data arriving from IFC exports are often imperfect. Material links can be incomplete, quantity fields inconsistent, and classifications can differ from one authoring application to another [19]. Work on semantic enrichment and ontology mapping has shown that schema records can be turned into machine-interpretable representations, but only when the mappings are explicit and the assumptions about information quality are stated [14,19]. For that reason, preprocessing in this study checks each entity for completeness, resolves its material name, and drops any record that fails the stated reliability rules. Figure 1 shows the path from the IFC file through to deterministic prompt packaging.

1.3. Gap in current approaches

The implementation responds to five limitations of current practice.

First, established LCA applications such as One Click LCA, Tally, and Athena Impact Estimator combine BIM quantities with environmental product declaration (EPD) data for whole-building assessment [20]. They are designed mainly for environmental accounting. Working through every combination of component materials in these tools during schematic design usually means iterating by hand, so they offer no direct counterpart to the discrete scenario search used in this study [20].

Second, multi-objective algorithms have been applied in building research to envelope parameters, operational energy, and life cycle cost [21–23]. Xue *et al.* [22] varied insulation thickness, window type, and window-to-wall ratio against cost and CO_2 emissions, and Kamazani and Dixit [24] considered embodied and operational carbon for envelope configurations. Such studies establish that multi-objective analysis is worthwhile, yet they usually address a single system or a continuous set of parameters. The question here is of a different kind: it concerns discrete material choices whose effects interact across walls, floors, and roofs.

Third, circularity measures seldom act as objectives in a design-space calculation. MCI, BCI, and related methods are informative [10,11], but they are generally applied for certification or material-passport purposes after design choices have been made. Using circularity as an objective requires proxy dimensions that can be attached to BIM components, such as recycled content, reuse potential, disassembly feasibility, and information completeness.

Fourth, decision support built on BIM has addressed sustainable component selection, for instance in the work of Jalaei *et al.* [25], who coupled BIM with a rule-based selection process. Pareto trade-offs were not examined in that work, and no component-aware queries were provided to explain why configurations differ on the carbon and circularity objectives. Most available outputs remain reports and do not offer a structured layer that can be interrogated.

Fifth, the use of LLMs in construction raises its own issue. Recent studies have examined generative artificial intelligence (AI)-assisted code-compliance checking in BIM and the architecture, engineering, and construction (AEC) industry as well as broader conversational-AI applications [26,27]. On their own, such applications provide neither a controlled interface to optimisation results derived from IFC nor deterministic guarantees. For this reason, the query and prompt-building stages developed here are rule-based, and any later interpretation by an LLM lies outside the system being evaluated.

1.4. Research objectives and scope

A deterministic Python prototype was developed to evaluate whole-building material scenarios. It runs in seven stages: IFC semantic extraction, data-quality filtering, material mapping, multi-objective enumeration, Pareto filtering, semantic querying, and prompt packaging for optional interpretation. The main elements are a direct STEP/SPF parser, a Building Circularity Score (BCS) built on four dimensions, 12 controlled query types, and structured prompts produced without any external API call.

The contributions fall into four areas:

(1) IFC-enabled semantic extraction. The prototype reads component geometry, quantities, and material information through IFC relationships. Explicit quality rules remove incomplete entities and record the reason for each exclusion, and the components that remain are mapped to material families ready for calculation.

(2) Whole-building formulation and exact enumeration. We treat the wall, floor, and roof materials as one coupled set of discrete decisions, minimising embodied carbon while maximising building circularity. With 150 configurations in the demonstrated space, every one of them can be evaluated

before Pareto filtering, and so exhaustive enumeration was used. For this space the answer is exact. We do not present the method as a stochastic or evolutionary optimiser.

(3) Controlled semantic queries. Twelve component-aware query types return defined patterns from the evaluated and non-dominated configurations. Each output is a reproducible summary drawn directly from the calculated fields.

(4) Deterministic prompt packaging. The Pareto topology, component choices, and trade-off descriptions are written into structured text that can be passed on for optional interpretation. No external API integration or model fine-tuning is needed for this step.

In combination, these elements treat the BIM model as a source of data for comparing materials, not merely as a visual or documentation model.

Scope statement. What is reported is a demonstrated deterministic prototype. Autonomous agents, OpenAI or Claude API calls, unrestricted natural-language parsing, Revit API integration, and knowledge-graph reasoning are outside its scope, as are cost optimisation, structural analysis, fire resistance, moisture behaviour, and other engineering feasibility constraints. Section 4 treats these as future extensions.

1.5. Paper organisation

The methods for extraction, scoring, enumeration, querying, and prompt packaging are set out in Section 2. Section 3 presents the IFC extraction, the non-dominated configurations and component interactions, the query outputs, and the robustness and runtime checks. Interpretation, limitations, and future development are discussed in Section 4, and Section 5 concludes the paper.

2. Methods

2.1. Overall framework

Seven deterministic stages make up the calculation: IFC extraction, data-quality filtering, material mapping, multi-objective enumeration, Pareto filtering, semantic querying, and prompt packaging (Figure 1). Every stage takes as input what the previous one produced. Component identities and quantities come from the IFC file, carbon and circularity parameters come from the material database, and the last stages arrange the calculated scenarios so that they can be inspected. Quantities are therefore never entered by hand, and no proprietary optimisation environment is needed.

The parser begins by following the standard IFC relationships, which give each relevant element its geometry, materials, quantities, and storey. Records that cannot be used are dropped at the quality check, and the remaining material names are matched to the prototype database by fixed mapping rules. After that, we evaluate all 150 feasible wall-floor-roof combinations (6 wall × 5 floor × 5 roof) for total embodied carbon and for building-level circularity. Handling the three groups together is deliberate, since their quantities interact. Pareto filtering then keeps the non-dominated configurations, the controlled queries pull out defined subsets or explanations, and the packaging step formats the results in case they are to be communicated further downstream.

Underneath all of this sits a quantified material inventory, still tied to the IFC entities it came from. No machine-learning inference is used anywhere, either in building it or in scoring the scenarios. With the same model and the same parameters, the outputs do not change.

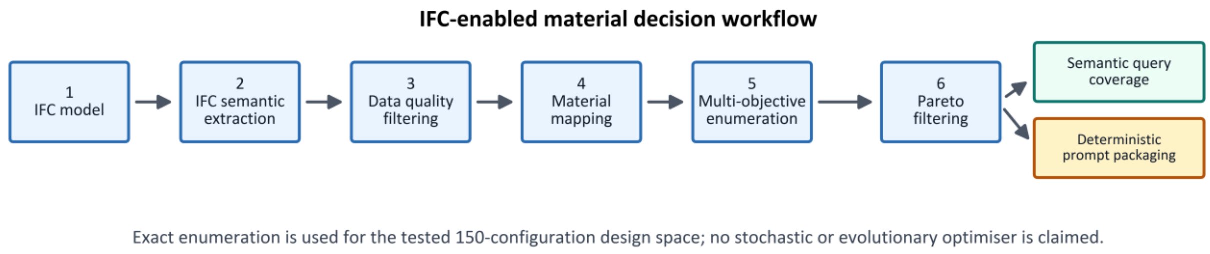

Figure 1. Workflow for IFC-based whole-building semantic decision support. Seven stages are shown. **(1)** The IFC model; **(2)** Semantic extraction from IFC; **(3)** Filtering for data quality; **(4)** Mapping of materials; **(5)** Multi-objective enumeration, in which material combinations are scored for embodied carbon and circularity; **(6)** Pareto filtering; **(7)** Semantic query coverage with deterministic prompt packaging. Enumeration is exact for the 150-configuration design space tested here, and no heuristic or evolutionary optimiser is claimed.

2.2. IFC semantic extraction

We limited extraction to the entity types that carry material in the envelope and structure of the reference file, namely IfcWall, IfcWallStandardCase, IfcSlab, and IfcRoof. IfcSlab records are split into floors and roofs when this is needed. The reason for the choice is practical. These elements hold quantities large enough to move both objectives, and IFC has defined routes from each of them to its material and quantity records.

Three IFC relationships give each element its context (Figure 2). We use IfcRelAssociatesMaterial to obtain the material or layer specification, IfcRelDefinesByProperties to reach the IfcElementQuantity records with volume or area measures, and IfcRelContainedInSpatialStructure to find the parent IfcBuildingStorey. All three are resolved for every element, so material identity, physical quantity, and location stay attached to the same component.

An element can carry several quantity measures at once. In that case the parser works down the list NetVolume > GrossVolume > Volume > NetArea > GrossArea > Area and takes the first field that has a value. Net volume heads the list because openings and voids have already been subtracted from it, so it comes closest to the actual amount of material.

Before an element is used in the calculation it has to pass three checks. Its material association must not be empty, its resolved quantity must be positive, and its entity type must be one we support. If any check fails, the record goes into an exclusion log with the reason. This happened once in the buildingSMART reference model, where an IfcRoof aggregate had no material and no usable quantity. We excluded it rather than invent a value for it (Table 1).

For parsing, we wrote our own pure-Python STEP/SPF text reader instead of relying on IfcOpenShell or a proprietary BIM API. It reads ISO 10303-21-style records into a dictionary of entities and walks the relationship paths above for IfcWall, IfcWallStandardCase, IfcSlab, and IfcRoof. Since it only handles text, there is no need for Revit, Dynamo, a compiled IfcOpenShell build, or any other proprietary runtime. The code runs on Python 3.10; the declared requirements are pandas ≥ 2.0.0, numpy ≥ 1.24.0, matplotlib ≥ 3.7.0, pytest ≥ 7.4.0, and streamlit ≥ 1.30.0.

Figure 2. Overview of IFC semantic extraction for the buildingSMART sample single-family house model. Top: counts of extracted elements (8), elements passing quality filtering (7), and excluded elements (1); Middle: table of extracted components with component ID, type, IFC material name, quantity and unit, and storey; Bottom left: the excluded component and the reason for its exclusion (IfcRoof aggregate container with no material name or quantity); Bottom right: checklist of the three validation rules used in filtering.

Table 1. IFC-extracted building components from the buildingSMART sample single-family house model (Building-Architecture.ifc). Seven valid components passed quality filtering: four walls (IfcWall), one floor slab (IfcSlab), and two roof slabs (IfcSlab with element_type = roof). Quantities extracted from IfcElementQuantity (NetVolume where available).

Component ID	Type	Element Name	Material	Quantity	Unit	Storey
WALL_001	wall	house-outer wall-house right front	stone_sand-lime	1.269	m^3	00 groundfloor
WALL_002	wall	house-outer wall-house right back	stone_sand-lime	1.786	m^3	00 groundfloor
WALL_003	wall	house-outer wall-house left	stone_sand-lime	4.231	m^3	00 groundfloor
WALL_004	wall	plumbing wall	gypsum_ fiber-board_panel	0.165	m^3	00 groundfloor
FLOOR_005	floor	floor	concrete_reinforced_ in-situ	6.438	m^3	00 groundfloor
ROOF_006	roof	house-roof-slab left	composite_element_roof	6.720	m^3	—
ROOF_007	roof	house-roof-slab right	composite_element_roof	9.364	m^3	—

2.3. Material mapping and data quality handling

Material names in IFC files are not written to any fixed taxonomy, so a database cannot simply look them up. The same family may turn up as “stone_sand-lime” or as “stone_sandlime”. Concrete labels may carry modifiers ("concrete_reinforced_in-situ", "concrete_precast"), and a name such as "composite_element_roof" says more about the build-up than about the material family. These variants had to be reconciled before any database parameters could be attached.

For this we use a keyword table. Its substring rules are checked in a fixed order, and edge cases fall through to a defined fallback. To give some examples, stone_sand-lime is mapped to masonry, gypsum_fiber-board to gypsum_board, concrete_reinforced (and the other concrete variants) to concrete, and composite_element_roof to roof_composite. No machine learning is involved. Because the rules are written out explicitly, a given label always receives the same mapping.

Put simply, each component goes through three steps here. It comes in carrying its source material string and resolved quantity. It is then tested for a material, a positive quantity, and a supported component type. Elements that fail are logged and taken out; those that pass move on with a mapped material family, which is what the carbon and circularity calculations need.

The prototype database holds 16 candidate materials spanning the wall, floor, and roof categories. For every option, Table S1 lists the density, cradle-to-gate carbon factor, four circularity dimensions, equal-weight BCS, source category, and verification notes. Part of the data is derived from databases, literature, or EPDs; the rest consists of explicitly stated prototype assumptions. The dataset is intended to demonstrate the calculation and is not presented as a verified, product-specific EPD database.

2.4. Building circularity representation

We use the BCS as an early-design proxy because the test case lacks the lifecycle information that MCI, BCI, or BECPI would need. The BCS should not be seen as a replacement for those fuller methods. Its job here is modest. It gives a transparent numerical objective that can be worked out for any material configuration from parameters linked to BIM components and stored in the prototype database.

Four proxy dimensions make up the BCS (Table 2). Each takes a value between 0 and 1, and a higher value means the material sits closer to the circular-economy attribute in question.

Table 2. Building Circularity Representation. The four BIM-extractable proxy dimensions from which the BCS is computed. In the prototype, each dimension is normalised to [0,1] and given the same weight ($w_i = 0.25$).

Dimension	Symbol	Description	Prototype Weight	Source
Recycled/renewable content	Crec	Fraction of material from recycled or renewable sources	0.25	EPD/ICE database/One Click LCA
Reuse/recyclability potential	Creu	Potential for reuse or recycling at end-of-life	0.25	Material property database
Disassembly potential	Cdis	Ease of disassembly for material recovery	0.25	Design attribute/product data
Information completeness	Cinf	Completeness and reliability of material data	0.25	Metadata/EPD verification level

Recycled or renewable content (C_{rec}) is the share of a material that comes from post-consumer recycled sources, pre-consumer industrial by-products, or rapidly renewable biological feedstocks. Reuse or recyclability potential (C_{reu}) rates how far a material can, technically and economically, be reused or recycled when its service life ends. Disassembly potential (C_{dis}) describes how easily a material or assembly can be physically separated from the building structure without being degraded. Information completeness (C_{inf}) reflects how much documentation is available to support the other three dimensions, and of what quality. In the prototype the four dimensions are weighted equally:

$$\text{BCSmaterial} = \Sigma(I = 1 \text{ to } 4)\ wi \cdot Ci \tag{1}$$

where w1 = w2 = w3 = w4 = 0.25. Equal weights were adopted because the framework is at prototype stage. They give a computationally tractable baseline for sensitivity testing and are not empirically validated circularity weights.

The influence of the weights was examined by setting the equal baseline (Crec = 0.25, Creu = 0.25, Cdis = 0.25, Cinf = 0.25) against four cases in which one dimension dominates: recycled content (0.40, 0.20, 0.20, 0.20), reuse/recyclability (0.20, 0.40, 0.20, 0.20), disassembly (0.20, 0.20, 0.40, 0.20), and information completeness (0.20, 0.20, 0.20, 0.40). For each case, the material scores, building circularity, all 150 configurations, Pareto filtering, and classification counts were recomputed. The exercise checks sensitivity within the prototype dataset; it does not calibrate the weights empirically.

At building level, circularity is the quantity-weighted aggregate of the component-level scores, so a material choice in a high-mass element has proportionally more influence on the overall score:

$$\text{BCbuilding} = \Sigma(\text{quantityj} \times \text{BCSj}) / \Sigma(\text{quantityj}) \tag{2}$$

2.5. Whole-building multi-objective enumeration and Pareto filtering

In this paper, optimisation refers to the stated two-objective decision problem and not to a heuristic search algorithm. The tested space contains only 150 discrete configurations, so evaluating all of them is both feasible and exact. Full enumeration removes the variation that stochastic search would introduce and ensures that no non-dominated configuration in the tested space is missed. A substantially larger design space would call for more scalable search or filtering under the same objective formulation, a point taken up in the runtime checks.

Wall, floor, and roof materials are the decision variables. They are evaluated jointly because the quantity assigned to any one component group shifts the whole building within the carbon-circularity plane. Figure 3 traces the calculation from IFC-derived quantities to the non-dominated material configurations.

2.5.1. Decision variables

Let c in {wall, floor, roof} denote the set of building component types extracted from the IFC model. For each component type c, let M_c denote the set of candidate material options drawn from the embedded material database: $|M_{wall}| = 6$, $|M_{floor}| = 5$, and $|M_{roof}| = 5$. A design option is defined as a material configuration tuple $m = (m_{wall}, m_{floor}, m_{roof})$ where m_c in M_c. The total decision space comprises all possible cross-system combinations:

$$|\text{Mwall}| \times |\text{Mfloor}| \times |\text{Mroof}| = 6 \times 5 \times 5 = 150 \tag{3}$$

A tuple thus defines one whole-building material scenario assembled from the available categories. A discrete representation suits the decision studied, which is a choice among candidate material systems and not the continuous adjustment of properties such as insulation thickness or density.

2.5.2. Embodied carbon objective

For each component i in the building model, the material mass is computed from the IFC-extracted geometric quantity and the density of the assigned material option:

$$mass_i = quantity_i \times \rho(m_c(i)) \tag{4}$$

where $\rho(m_c(i))$ denotes the density associated with material option m for component i of type c. The embodied carbon contribution of component i is then:

$$EC_i = mass_i \times CF(m_c(i)) \tag{5}$$

where CF(mc(i)) denotes the cradle-to-gate carbon factor (kilograms of carbon dioxide equivalent per kilogram, $kgCO_2e/kg$) of the assigned material. The total embodied carbon objective aggregates across all N components in the extracted building model:

$$TotalEC(m) = \Sigma(i=1 \text{ to } N)\ EC_i \tag{6}$$

2.5.3. Building circularity objective

The circularity objective operates at the building level through quantity-weighted aggregation of material-level BCS (Section 2.4). For a given material configuration m, the building-level circularity is computed as:

$$BuildingCircularity(m) = \Sigma(quantity_i \times BCS(m_c(i))) / \Sigma(quantity_i) \tag{7}$$

The quantity weighting gives materials used in larger volumes a correspondingly greater influence on the building-level score.

2.5.4. Pareto formulation

The multi-objective decision formulation simultaneously minimises total embodied carbon and maximises building-level circularity:

$$\min_m TotalEC(m);\ \max_m BuildingCircularity(m) \tag{8}$$

Pareto dominance is defined as follows: a material configuration m_A strictly dominates m_B if and only if:

$$TotalEC(m_A) \leq TotalEC(m_B) \text{ AND } BuildingCircularity(m_A) \geq BuildingCircularity(m_B) \tag{9}$$

with at least one inequality strict. The Pareto-optimal set comprises all configurations for which no dominating alternative exists:

$$P = \{m \mid \text{no } m' \text{ dominates } m\} \tag{10}$$

All 150 combinations are evaluated and then compared pairwise for dominance. The procedure is deterministic and is not an evolutionary optimiser. It identified 12 configurations (8%) that are Pareto-optimal in the formal sense that none is dominated by another evaluated configuration.

Runtime was measured for the 150-configuration case and for synthetic sets of 1000, 10,000, and 100,000 configurations. Pairwise $O(n^2)$ filtering was retained for 150 and 1000 cases. At 10,000 and 100,000 configurations it was omitted as impractical, and a sorted two-objective filter was timed instead.

Joint evaluation is required because Pareto status depends on combinations across component types. In this test case, the larger floor and roof quantities create stronger trade-offs than the wall quantities, an interaction that would be hidden by separate single-component comparisons (Figure 3).

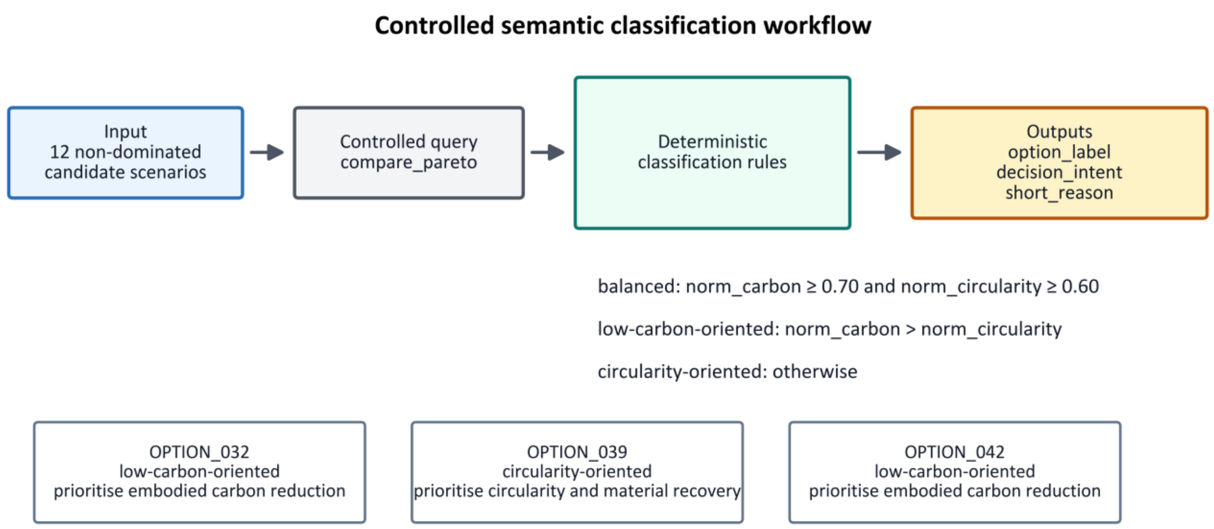

Figure 3. Whole-building enumeration and controlled semantic query workflow. The diagram shows how the compare_pareto query classifies 12 non-dominated candidate scenarios using deterministic rules based on $norm_{carbon}$ and $norm_{circularity}$. The output fields are option_label, decision_intent, and short_reason, illustrated with selected candidate cards rather than a dense table.

2.5.5. Semantic classification

Classification places each non-dominated configuration according to its relative position between the two objectives. As a first step, carbon and circularity values are normalised to the ranges observed across all 150 combinations:

$$\text{normcarbon} = (\max(\text{TotalEC}) - \text{TotalEC}) / (\max(\text{TotalEC}) - \min(\text{TotalEC})) \tag{11}$$

$$\text{normcircularity} = (\text{BuildingCircularity} - \min(\text{BC})) / (\max(\text{BC}) - \min(\text{BC})) \tag{12}$$

Three rule-based labels are assigned from the normalised values. A configuration is labelled balanced if normcarbon $\geq$ 0.70 and normcircularity $\geq$ 0.60. If it does not meet both balanced thresholds, it is low-carbon-oriented where normcarbon > normcircularity and circularity-oriented where normcircularity $\geq$ normcarbon. The labels describe which objective an evaluated configuration leans towards; they do not introduce a further optimisation step.

Threshold sensitivity was examined on a 3 × 3 grid that combined normcarbon thresholds of 0.60, 0.70, and 0.80 with normcircularity thresholds of 0.50, 0.60, and 0.70. The grid is meant to show whether labels shift within this design space; it does not establish thresholds for general use.

2.5.6. Scenario-based uncertainty propagation

Because the database holds point estimates and not empirically calibrated probability distributions, uncertainty propagation was set up as a scenario-based stress test. Carbon factors were sampled from triangular distributions centred on their baseline values with ±20% bounds. Each of the four BCS dimensions was sampled from a triangular distribution with ±0.10 absolute bounds and then clipped to [0,1]. The analysis ran for 1000 iterations (random seed = 42); in each iteration, all 150 configurations and their Pareto set were recalculated, and the retention frequency of every baseline candidate was recorded. Two further cases, one perturbing carbon factors only and one perturbing BCS dimensions only, separate the two sources of uncertainty, and their Pareto sets were compared with the baseline using mean Jaccard similarity. The chosen ranges are transparent assumptions made for this prototype, not product-specific EPD uncertainty intervals.

2.6. Semantic query layer

The query layer takes as input the 150 calculated configurations, with their objective values and Pareto flags. Instead of a conversational interface, it offers twelve predefined queries, each governed by fixed retrieval and explanation rules, so no natural-language processing, machine-learning inference, or probabilistic reasoning influences a response. This makes the outputs reproducible, although it does not test whether designers find them useful.

The twelve queries form four groups. The global queries low_carbon and high_circularity rank all 150 options on a single objective, and balanced returns the configurations that satisfy both thresholds. Six component queries pick out the best-performing wall, floor, or roof material on either the carbon or the circularity criterion. pareto_only and compare_pareto work on the 12 non-dominated configurations, and compare_pareto also assigns the deterministic labels and decision intents. Lastly, explain_option returns a structured trace for one selected configuration. The scope and output of every query are listed in Table 3.

Coverage was checked against five information dimensions. Carbon ranking is served by low_carbon together with the three component carbon queries, and circularity ranking by high_circularity together with its three component queries. Component attribution draws on those six component queries and explain_option; Pareto status is handled by pareto_only and compare_pareto; and threshold classification by balanced and compare_pareto. Formally, for I = {carbon ranking, circularity ranking, component attribution, Pareto status, threshold classification}, every dimension is addressed by at least one member of $Q = \{q1,...,q12\}$. The repository tests cover ranking direction, normalisation, component-field consistency, Pareto filtering, known threshold classifications, invalid queries, and export/read-back behaviour. Passing them shows that the rules behave correctly for this case; it is not empirical evidence of designer interpretability.

Table 3. Semantic query taxonomy: twelve controlled query types organised by functional scope. Each query maps to deterministic retrieval and explanation logic over the pre-computed optimisation results.

Query Type	Scope	Description	Output
low_carbon	Global	All material configurations sorted by ascending embodied carbon	Ranked list of 150 options with carbon values
high_circularity	Global	All material configurations sorted by descending circularity score	Ranked list of 150 options with circularity scores
balanced	Global	Configurations meeting both normalised threshold criteria	Filtered subset with dual-threshold compliance
low_carbon_wall	Component	Wall material option yielding the lowest embodied carbon	lowest-carbon wall material within the tested design space
low_carbon_floor	Component	Floor material option yielding the lowest embodied carbon	lowest-carbon floor material within the tested design space
low_carbon_roof	Component	Roof material option yielding the lowest embodied carbon	lowest-carbon roof material within the tested design space
high_circularity_wall	Component	Wall material option yielding the highest circularity score	highest-circularity wall material within the tested design space
high_circularity_floor	Component	Floor material option yielding the highest circularity score	highest-circularity floor material within the tested design space
high_circularity_roof	Component	Roof material option yielding the highest circularity score	highest-circularity roof material within the tested design space
pareto_only	Pareto	The set of 12 Pareto-optimal configurations	Non-dominated subset of the design space
compare_pareto	Pareto	All 12 Pareto-optimal configurations annotated with semantic classification labels	Fully classified Pareto set with natural-language explanations
explain_option	Special	Structured reasoning trace for a specified material configuration	Component-aware explanation of the option's Pareto position

2.7. Deterministic prompt-packaging workflow

Prompt packaging is kept apart from query execution and from any LLM inference. The code writes the selected optimisation results into a standard template, which a user may choose to transfer to an LLM interface. No external model is called, and neither model reasoning nor designer usability is assessed.

Six sections make up the template, and compare_pareto fills all of them. Section one gives the IFC component count and defines the objectives. Sections two and three group the 12 non-dominated configurations by label and list their component materials and objective values. Section four asks for a summary of what the candidates imply (explicitly not a design prescription). Section five describes the Pareto trade-off, and section six sets out the limits of the method: the envelope-material scope, the cradle-to-gate carbon factors, and the four-dimension BCS.

Because every field is taken from a single query result, re-running the step on unchanged data gives word-for-word the same text. We did not evaluate LLM outputs, compare models, or choose an LLM service; these questions sit outside the study.

2.8. Streamlit prototype

We built a seven-page Streamlit interface so that the whole process could be exercised in one place (Figure 4). The case is introduced on the Overview page, and the IFC file is loaded on IFC Upload. Configurations are evaluated on Optimisation, and the non-dominated trade-offs can be inspected on

Pareto Analysis. Query Engine runs the controlled queries, LLM Prompt puts together the prompt text, and Export writes out the reproducibility outputs.

No separate backend service is needed, because the interface imports the calculation, query, and interpretation modules directly. Selections made on one page are kept in session state for the next. The interface should be read as a research demonstrator; it is not a production BIM platform.

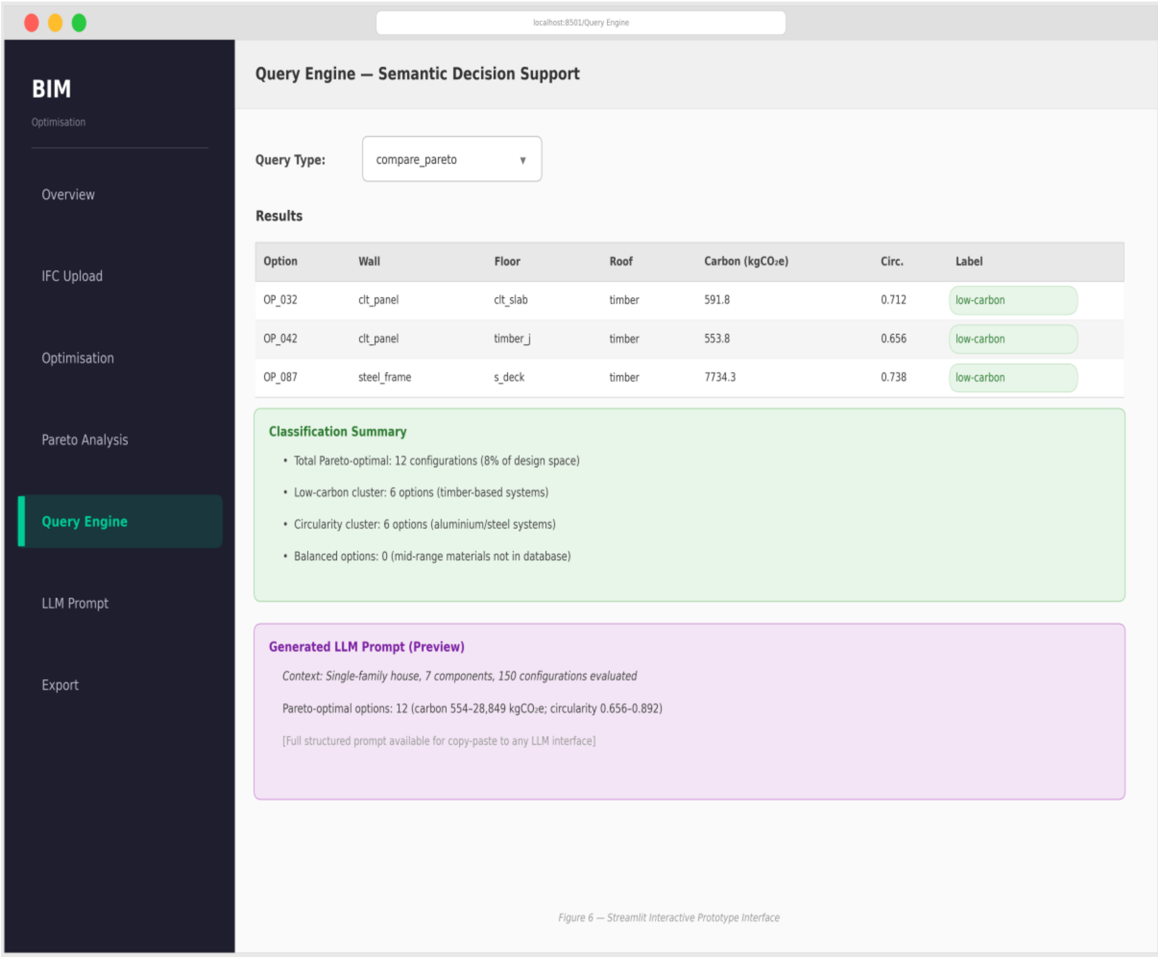

Figure 4. Query Engine page of the interactive Streamlit prototype. The sidebar on the left links the seven pages (Overview, IFC Upload, Optimisation, Pareto Analysis, Query Engine, LLM Prompt, and Export). The main panel holds the query-type selector, a table of classification results, a classification summary, and a preview of the generated LLM prompt.

3. Results

3.1. IFC extraction results

For the extraction test we used Building-Architecture.ifc, a reference model from the buildingSMART Sample-Test-Files. The parser found eight elements of the target types (IfcWall, IfcSlab, and IfcRoof), which then went through the three-pathway semantic retrieval and the data-quality filter set out in Section 2.2 and Section 2.3.

Of these eight, seven passed every validation rule and went on to optimisation. The eighth was excluded, with the reason logged. What remained were four exterior walls, one ground-floor slab, and two roof slabs (Table 1, Figure 2). Quantities varied widely, from 0.165 m³ for the plumbing wall

(WALL_004, gypsum fibre-board panel) up to 9.364 m^3 for the right roof slab (ROOF_007, composite element roof). The walls came to 7.451 m^3 in total, the floor slab to 6.438 m^3, and the two roof slabs to 16.084 m^3, giving 30.039 m^3 overall. It is this total that provides the quantity weighting for the building-level circularity results in Section 3.2.

The excluded element was an IfcRoof aggregate instance. It had no material association and no quantity that could be resolved, so the pipeline logged the reason and left it out of the optimisation set. Anyone checking the log can see which IFC schema gap caused the exclusion; the element was not dropped silently.

In short, the reference model was converted into a representation the optimiser could use, and its semantic context and data-quality record came through intact. It is worth noting that even a reference-quality file contained an incomplete entity. The validation rules are therefore needed for curated IFC models too, since without explicit filtering a gap of this kind would pass errors unnoticed into the optimisation calculations.

3.2. Optimisation results

Walls (4 components), floor (1 component), and roof (2 components) were the three building systems into which the seven extracted components were grouped, and their candidate materials came from the database described in Section 2.3 and Section 2.4. Every combination was generated, giving a design space of 150 whole-building material configurations (6 wall × 5 floor × 5 roof); total embodied carbon and building-level circularity were then calculated for each using the formulations in Section 2.5.

Of the 150 configurations, 12 (8%) were Pareto-optimal in the formal non-dominated sense (Figure 5); for none of these candidate scenarios does any alternative improve both objectives at once. The 12 do not trace a smooth, continuous frontier but fall into two separate groups. One occupies the low-carbon region (553.8–7,735 $kgCO_2e$, circularity 0.656–0.738) and the other the high-circularity region (21,684–28,849 $kgCO_2e$, circularity 0.810–0.892). The empty interval between them indicates that the material database has no intermediate options able to cover the mid-range of the trade-off, a point examined further in Section 3.3.

Objective values across the full design space spanned 553.8 to 36,811.3 $kgCO_2e$ for embodied carbon and 0.350 to 0.895 for building circularity (Figure 6). The lower bound corresponds to a configuration with Cross-Laminated Timber (CLT) panel walls, timber joist floor, and timber roof; the upper bound corresponds to aerated concrete walls, in-situ reinforced concrete floor, and aluminium roof. This extreme range reflects the intentionally contrasting prototype material database rather than representative market distributions; it illustrates the sensitivity of whole-building embodied carbon to material-system coupling under maximally divergent material choices.

Enumeration thus reduced a structured design space to a non-dominated set of whole-building candidate scenarios that can be interpreted directly. Since all 150 configurations were evaluated, with no heuristic sampling, no non-dominated candidate scenario within the demonstrated design space can have been overlooked.

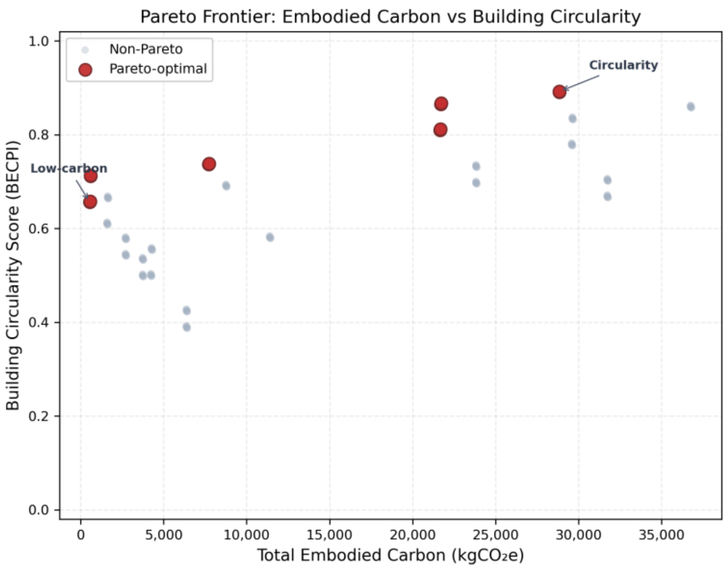

Figure 5. Pareto trade-off frontier showing embodied carbon versus building circularity score for 150 material configurations. Pareto-optimal configurations (n = 12) highlighted in red. Non-Pareto configurations shown as translucent grey circles.

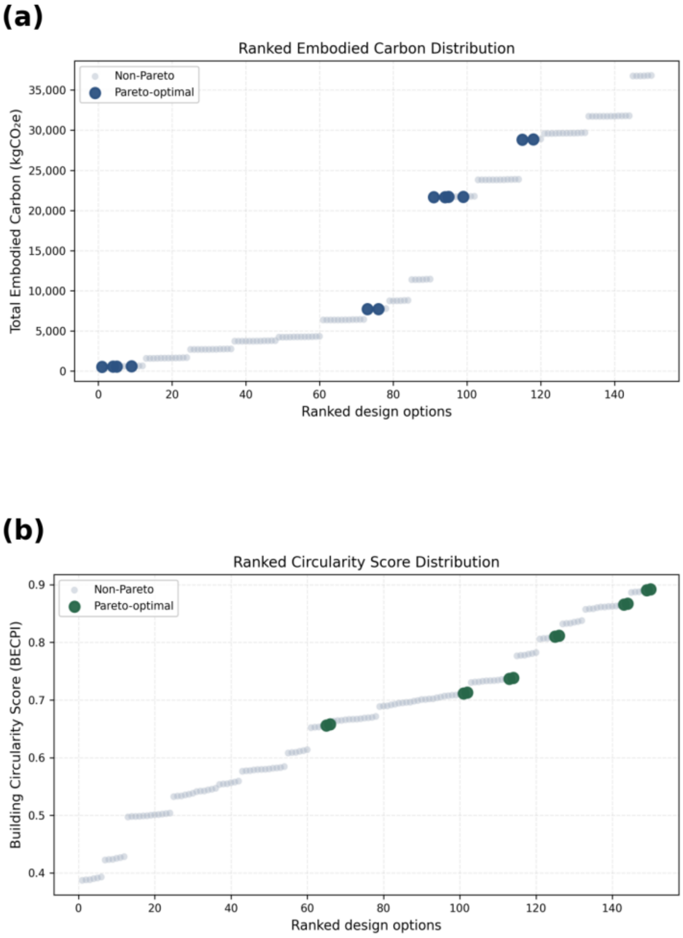

Figure 6. Ranked distribution of embodied carbon **(a)** and building circularity score **(b)** across all 150 material configurations. Options sorted by ascending carbon **(a)** and descending circularity **(b)**. Pareto-optimal configurations highlighted. Carbon range: 553.8–36,811.3 $kgCO_2e$. Circularity range: 0.350–0.895.

3.3. Whole-building trade-off behaviour

This subsection uses the non-dominated candidate scenarios to ask whether material choices become decision-relevant when wall, floor, and roof are evaluated as a coupled configuration instead of as isolated component substitutions.

Cluster A—Low-carbon-oriented configurations. Six of the non-dominated candidate scenarios belong to this group (Table 4). Each uses timber_roof, combined with CLT panel or steel frame walls and one of three floor materials (CLT slab, steel beam deck, or timber joist). Embodied carbon runs from 553.8 $kgCO_2e$ for OPTION_042 (CLT panel, timber joist, timber roof) to 7734.3 $kgCO_2e$ for OPTION_087 (steel frame, steel beam deck, timber roof), and building circularity from 0.656 to 0.738. Floor material accounts for the largest variation inside the cluster. Replacing timber joist with steel beam deck adds approximately 7,165 $kgCO_2e$ while raising circularity by only 0.081, because the floor slab volume is large (6.438 m^3) and steel has a high carbon factor.

Cluster B—Circularity-oriented configurations. The remaining six candidates have the same wall and floor pairings as Cluster A, but their roof is aluminium_roof. Carbon in this cluster ranges from 21,668.5 $kgCO_2e$ (OPTION_044) to 28,849.0 $kgCO_2e$ (OPTION_089), and circularity from 0.810 to 0.892. Because aluminium is recyclable at end of life, the aluminium roof pushes circularity above 0.80 in every case. The price is a substantially higher carbon figure than anywhere in Cluster A.

Roof-driven trade-off. Pairs that share walls and floor but differ in roof material make the roof's influence plain. With CLT panel walls, a CLT slab, and a timber roof, OPTION_032 comes to 591.8 $kgCO_2e$ at 0.712 circularity. With an aluminium roof instead (OPTION_034), the figures become 21,706.5 $kgCO_2e$ and 0.865, so carbon goes up 37-fold while circularity rises by 22%. The steel frame and timber joist pair behaves in the same way. OPTION_092, with a timber roof, gives 569.4 $kgCO_2e$ and 0.658; OPTION_094, with aluminium, gives 21,684.1 $kgCO_2e$ and 0.812, a 38-fold carbon increase for 23% more circularity. The roof decides the trade-off mostly because the two roof slabs hold 53.5% of all the material volume we extracted.

Absence of balanced options. No Pareto-optimal option met both normalised thresholds ($norm_{carbon} \geq 0.70$ and $norm_{circularity} \geq 0.60$), so there are zero balanced configurations. We read this as a consequence of what is in the material database, and not as a methodological failure. Nothing in the prototype database can sit between the two clusters; it has no hybrid timber–steel roof systems, for example, and no recycled-content structural materials.

Component interaction hierarchy. In both clusters, floor and roof choices move the results far more than wall choices do. Once the roof is fixed, a given wall material lands at similar carbon values. CLT panel walls, for instance, give 553.8–591.8 $kgCO_2e$ with the timber roof and 21,668.5–21,706.5 $kgCO_2e$ with the aluminium roof. The four walls make up only 24.8% of the extracted volume, which is why the wall material has so little leverage in this prototype case. The trade-off comes out of how the systems are combined; no single material produces it on its own.

For the case we tested, then, the Pareto frontier picks up effects that only appear at whole-building level. Material choices become relevant to the decision only when walls, floors, and roofs are assessed together as one configuration; swapping components one at a time does not reveal this. We regard it as an observation about the prototype case and make no claim that it applies more generally.

3.4. Semantic query results

The semantic query layer was exercised by executing the twelve controlled query types defined in Section 2.6 against the 150 evaluated configurations. The compare_pareto query classified all 12 non-dominated candidate scenarios, returning 6 low-carbon-oriented, 6 circularity-oriented, and 0 balanced configurations (Table 4). Each option received a decision_intent field and a short_reason grounded in the underlying objective values. Low-carbon-oriented options were annotated with the intent 'prioritise embodied carbon reduction'; circularity-oriented options with 'prioritise circularity and material recovery.' The absence of balanced options was reported explicitly, confirming deterministic query operation.

Component-specific queries decomposed whole-building results into per-component material drivers. The low_carbon_wall query identified CLT panel as the wall material yielding the lowest embodied carbon within the tested 150-configuration design space. The high_circularity_roof query identified aluminium roof as the roof material yielding the highest circularity score within the tested 150-configuration design space. These per-component results confirmed that the semantic classification at the whole-building level aligns with component-level rankings: every low-carbon-oriented Pareto option employs timber roof, while every circularity-oriented option employs aluminium roof.

The query layer converts numeric optimisation output into component-aware semantic explanation without employing natural language processing or machine learning. Each response is fully deterministic: identical queries against identical input data produce identical classifications and explanations. Query correctness was checked by comparing deterministic outputs against the underlying sorted objective values, component fields, Pareto flags, and threshold rules used to generate the 150-configuration design space. This result demonstrates deterministic query coverage, not validated designer interpretability or usefulness.

Table 4. Non-dominated whole-building material configurations (12 of 150 options). Each row represents one non-dominated design option with the selected wall, floor, and roof material; total embodied carbon ($kgCO_2e$); building circularity score; and semantic classification label. The low-carbon-oriented cluster (6 options) is dominated by CLT/timber systems with carbon below 7800 $kgCO_2e$. The circularity-oriented cluster (6 options) is characterised by aluminium_roof systems and circularity scores ranging from 0.810 to 0.892. No option satisfies the balanced threshold.

Option ID	Wall	Floor	Roof	Carbon ($kgCO_2e$)	Circularity	Label
OPTION_032	clt_panel	clt_slab	timber_roof	591.8	0.712	low-carbon-oriented
OPTION_034	clt_panel	clt_slab	aluminium_roof	21,706.5	0.865	circularity-oriented
OPTION_037	clt_panel	steel_beam_deck	timber_roof	7718.7	0.737	low-carbon-oriented
OPTION_039	clt_panel	steel_beam_deck	aluminium_roof	28,833.4	0.891	circularity-oriented
OPTION_042	clt_panel	timber_joist	timber_roof	553.8	0.656	low-carbon-oriented
OPTION_044	clt_panel	timber_joist	aluminium_roof	21,668.5	0.810	circularity-oriented
OPTION_082	steel_frame	clt_slab	timber_roof	607.3	0.713	low-carbon-oriented
OPTION_084	steel_frame	clt_slab	aluminium_roof	21,722.0	0.867	circularity-oriented
OPTION_087	steel_frame	steel_beam_deck	timber_roof	7734.3	0.738	low-carbon-oriented
OPTION_089	steel_frame	steel_beam_deck	aluminium_roof	28,849.0	0.892	circularity-oriented
OPTION_092	steel_frame	timber_joist	timber_roof	569.4	0.658	low-carbon-oriented
OPTION_094	steel_frame	timber_joist	aluminium_roof	21,684.1	0.812	circularity-oriented

Table 5 summarises the deterministic semantic classification outputs for all 12 non-dominated candidate scenarios, including the assigned label, decision intent, and rule-grounded explanation generated by the compare_pareto query.

Table 5. Semantic query classification results for the 12 non-dominated candidate scenarios using the compare_pareto query. Each option is classified by its relative position on the normalised carbon-circularity plane and annotated with a decision intent and natural-language explanation.

Option ID	Label	Decision Intent	Reason
OPTION_032	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_034	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.
OPTION_037	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_039	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.
OPTION_042	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_044	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.
OPTION_082	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_084	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.
OPTION_087	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_089	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.
OPTION_092	low-carbon-oriented	prioritise embodied carbon reduction	This option leans toward low embodied carbon while remaining non-dominated.
OPTION_094	circularity-oriented	prioritise circularity and material recovery	This option leans toward high circularity while remaining non-dominated.

3.5. Deterministic prompt packaging and prototype demonstration

To check the prompt packaging, we passed the compare_pareto results into the six-section template from Section 2.7. The resulting prompt had six semantic sections: context, the grouped non-dominated candidate configurations, a full tabular listing, a request to interpret the candidate scenarios, a trade-off summary, and the limitations. Each section is filled deterministically from the query results, and running the step again gives an identical prompt. The step makes no external API calls, and we did not evaluate any LLM output.

Nothing in the process generates design prescriptions autonomously. What it produces is context already structured for a prompt, capturing the topology of the Pareto front, the semantic labels, and the component-level material drivers as summaries of candidate scenarios that a user may choose to interpret later. Our checks covered prompt structure and content completeness. LLM inference quality, user trust, and recommendation accuracy were outside them.

The seven-page Streamlit prototype (Figure 4) runs the whole workflow through its Overview, IFC Upload, Optimisation, Pareto Analysis, Query Engine, LLM Prompt, and Export pages. It imports the optimisation, query, and interpretation modules directly, so there is no separate backend service. Running the full pipeline in this way confirmed that the parts work together as a single reproducible system. We offer it as a research demonstrator rather than a production platform.

Several limitations apply to these results. Only one IFC reference file was tested, and it gave seven valid components. The material database is a sample of 16 options over three component types; it makes no claim to cover the market or to be verified against product-specific EPDs. Circularity is scored from proxy dimensions under tested weighting schemes, and the weights have not been calibrated empirically. The carbon factors are cradle-to-gate, so end-of-life and biogenic carbon dynamics are left out. These points limit what the prototype verification covers; they do not reduce the contribution of the workflow as demonstrated.

3.6. Sensitivity, reproducibility and runtime checks

Table S1 reports all 16 prototype material options used in the workflow and includes each material's density, carbon factor, circularity dimensions, equal-weight BCS, source category, and notes. The table distinguishes values classified as database/literature/EPD-derived from values classified as prototype demonstration assumptions.

Across all five BCS weighting schemes, the non-dominated candidate set remained at 12 configurations with 12/12 overlap relative to the equal-weight baseline. Classification remained 0 balanced, 6 low-carbon-oriented, and 6 circularity-oriented candidate scenarios. The two-cluster structure persisted under all tested weighting schemes. This supports robustness within the demonstrated prototype case only.

Detailed scenario-level results for the five BCS weighting schemes are reported in Table S2.

Across the nine tested threshold combinations, classification counts remained unchanged relative to the baseline: 0 balanced, 6 low-carbon-oriented, and 6 circularity-oriented candidate scenarios. Therefore, the absence of balanced configurations in this case is not solely an artefact of the selected 0.70/0.60 baseline threshold. This demonstrates classification stability only for the tested design space and does not establish a universal threshold rule.

Classification outcomes for all nine threshold combinations are provided in Table S3.

The current 150-configuration case completed in approximately 0.79 s, with pairwise $O(n^2)$ Pareto filtering requiring approximately 0.023 s. For 1000 synthetic configurations, pairwise Pareto filtering required approximately 0.386 s. For 10,000 and 100,000 synthetic configurations, pairwise filtering was skipped to avoid impractical runtime, and sorted two-objective Pareto filtering required approximately

2.27 s and 32.67 s respectively. Exhaustive enumeration is practical for the demonstrated small design space, while larger spaces require more scalable Pareto-filtering strategies.

Runtime measurements across the tested design-space sizes are reported in Table S4.

3.7. Scenario-based uncertainty propagation

Monte Carlo uncertainty propagation was performed using 1000 sampled runs over the same 150-configuration design space. Under the combined uncertainty scenario (carbon factors ±20%; BCS proxy dimensions ±0.10), all 12 baseline non-dominated candidate scenarios remained non-dominated with frequencies of at least 0.832. Four configurations remained non-dominated in every sampled run: OPTION_032 (CLT panel, CLT slab, timber roof), OPTION_082 (steel frame, CLT slab, timber roof), OPTION_034 (CLT panel, CLT slab, aluminium roof), and OPTION_084 (steel frame, CLT slab, aluminium roof). The lowest-frequency baseline scenarios were OPTION_092 (0.833) and OPTION_094 (0.832), both involving steel frame walls and timber joist floors with timber or aluminium roofs.

Across the 1000 sampled runs, the non-dominated set size ranged from 4 to 24 configurations (mean 11.945, standard deviation 1.728), indicating that uncertainty occasionally expands or contracts the candidate decision space but does not systematically inflate it far beyond the baseline 12 configurations. New non-dominated configurations outside the baseline Pareto set appeared in 8.7% of sampled runs, suggesting that most uncertainty effects involved stability changes within or near the baseline candidate set rather than frequent emergence of wholly new trade-off configurations. The source-specific uncertainty comparison indicated that carbon-factor perturbation had a larger effect on Pareto-set stability than BCS-dimension perturbation under the tested assumptions. The mean Jaccard similarity between sampled and baseline non-dominated sets was 0.951 for carbon-only uncertainty and 0.992 for BCS-only uncertainty. This suggests that, for the demonstrated material database and quantity structure, embodied-carbon factors are the more influential uncertainty source for whether an option remains non-dominated. This result is an uncertainty stress test within the prototype case and should not be interpreted as a general sensitivity ranking across material databases or building typologies.

The full scenario-level Pareto-retention frequencies are provided in Table S5, and the aggregate uncertainty statistics are reported in Table S6.

4. Discussion

4.1. Prototype scope and limitations

What do the robustness checks change? Chiefly, they show that the two-cluster Pareto structure does not hinge on one particular setting. It held under five BCS weighting schemes, nine combinations of classification thresholds, and the Monte Carlo perturbations stated earlier. Each of these probes something different (the circularity weights we chose, the labelling rules, and the input parameters), so their agreement carries some weight, but only for this dataset and this IFC case. Nothing here shows robustness across building types, material markets, or empirically calibrated circularity models. The split into two clusters is itself a product of the current materials. The roof slabs account for 53.5% of extracted volume, and the database has no mid-range options, such as hybrid timber–steel roof systems or recycled-content structural materials, to bridge the gap. Much of what practical specification would need is also missing, including cost and procurement, structural and fire performance, moisture and durability, operational energy, end-of-life and biogenic carbon, code compliance, live model synchronisation, formal compatibility knowledge, and user evaluation. We did not test LLM inference or the behaviour of autonomous agents.

The evidence is narrow. It rests on a single IFC reference file for one building type, with seven accepted components and 16 prototype material options. Getting the pipeline to operate on that file verifies the prototype, but it does not validate the approach across authoring systems or IFC export conventions. Models from Revit, Archicad, Tekla, and other platforms would have to be tested before anything could be said about the robustness of the extraction or about field applicability.

Within those limits, the study does verify several things. STEP-formatted IFC records can be read and filtered while a log is kept of what was excluded. Material names can be mapped with deterministic rules. Every whole-building combination can be enumerated for the two objectives we set. The non-dominated results can then be queried, classified, packaged, and looked at in Streamlit. Linked together, these steps make up the proof of concept. They cannot tell us whether the same performance would be achieved beyond the reference case.

4.2. Future directions

Turning this demonstrator into an operational design application would involve three distinct stages.

The near-term stage concerns the evidence base and the feasible design space. Cost, structural performance, fire resistance, moisture behaviour, and related constraints could be added as objectives or as filters. Introducing verified EPD and lifecycle data, hybrid systems, and recycled-content materials would show whether the low-carbon and circularity clusters remain separate once intermediate options are available. The component-leverage pattern also cannot be generalised beyond the reference house until commercial, multi-storey residential, and other IFC models have been tested.

In a second stage, material comparison would be linked to explicit feasibility knowledge. Compatibility rules for structure, moisture, thermal expansion, fire performance, and regulatory constraints could be represented in a formal knowledge graph in place of the current keyword mapping. Interaction assisted by language models should be evaluated only after suitable guardrails and protocols have been defined, since this study tests prompt packaging and not LLM reasoning.

Further ahead, the calculation could be scaled up to portfolios of buildings and tried out with practising designers. Such a study would have to look at how designers interpret the outputs, how confident they are in their decisions, and whether they would adopt the approach under realistic project conditions. None of this is covered by the present proof of concept.

5. Conclusion

A geometric model on its own is not enough for comparing materials across a whole building. The quantities of each component and the properties of its material have to stay linked while carbon is weighed against circularity. In the process we have described, those data are read from IFC, discrete wall-floor-roof scenarios are calculated, and the non-dominated results are organised with controlled queries and deterministic prompt packaging.

The implementation makes four contributions. Direct IFC extraction recovered seven valid components from a buildingSMART reference model, with the reasons for excluded records logged. Exact enumeration of 150 whole-building configurations left 12 non-dominated carbon-circularity candidates. The 12 controlled query types returned reproducible, component-aware outputs, among them a classification of 0 balanced, 6 low-carbon-oriented, and 6 circularity-oriented scenarios. Finally, prompt packaging rendered the same calculated records as structured text without an external API call.

Changing between the five BCS weighting schemes left the candidate set unchanged, and changing between the nine threshold combinations left the classification counts unchanged. Exhaustive enumeration proved lightweight for the 150 tested configurations, although larger synthetic sets require more scalable Pareto filtering. These checks make the reference-case interpretation more secure but do not extend it to other buildings or material databases.

What the study provides, then, is prototype-stage evidence. The analysis did not cover structural performance, fire resistance, moisture behaviour, durability, cost, procurement, or regulatory compliance. Field use could be considered only once the process has been run on multiple IFC models, supplied with verified EPD and lifecycle data, and constrained by engineering feasibility, and once both user interpretation and LLM-output quality have been evaluated.

Data availability statement

The data and code supporting the findings of this study are openly available in the project GitHub repository at https://github.com/Yiping0122/bim-material-decision-tool. Supplementary files are supplied with the manuscript: Table S1 contains the full prototype material database; Table S2 reports BCS weighting sensitivity; Table S3 reports semantic-classification threshold sensitivity; Table S4 reports runtime and scalability checks; Table S5 reports scenario-level Monte Carlo Pareto-retention frequencies; and Table S6 reports aggregate non-dominated-set size, Jaccard-similarity, and new-solution statistics. The prototype was implemented in Python 3.10 using an in-house pure-Python STEP/SPF IFC parser and standard open-source libraries. The repository requirements specify pandas ≥ 2.0.0, numpy ≥ 1.24.0, matplotlib ≥ 3.7.0, pytest ≥ 7.4.0, and streamlit ≥ 1.30.0. The repository additionally contains the Python implementation, requirements, unit tests, revision outputs, generated tables, and scripts required to reproduce the demonstrated workflow and associated sensitivity, scalability, and uncertainty analyses.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used Grammarly to assist with language editing and formatting. The tool was not involved in the research design, data analysis, or interpretation of results. The authors reviewed and edited all AI-generated content.

Authors' contribution

Conceptualization, Y.M. and H.H.; methodology, Y.M.; software, Y.M. and Y.S.; formal analysis, Y.M.; writing—original draft preparation, Y.M.; writing—review and editing, Y.S., H.H. and C.Z.; visualization, Y.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Röck M, Saade MR, Balouktsi M, Rasmussen FN, Birgisdóttir H, *et al*. Embodied GHG emissions of buildings—the hidden challenge for effective climate change mitigation. *Appl. Energy* 2020, 258:114107.
[2] GlobalABC. Global Status report for buildings and construction 2025–2026. 2026. Available: https://globalabc.org/global-status-report (accessed on 10 August 2026).
[3] International Energy Agency (IEA). Net zero by 2050. 2021. Available: https://www.iea.org/reports/net-zero-by-2050 (accessed on 10 August 2026).
[4] European Committee for Standardization (CEN). EN 15978:2026 sustainability of construction works—assessment of environmental performance of buildings—requirements and guidance. 2026. Available: https://www.cencenelec.eu/news-events/news/2026/en-in-the-spotlight/2026-04-17-en-15978-2026/ (accessed on 10 August 2026).
[5] International Organization for Standardization (ISO). ISO 21931-1:2022. Sustainability in buildings and civil engineering works—framework for methods of assessment of the environmental, social and economic performance of construction works as a basis for sustainability assessment—part 1: buildings. 2022. Available: https://www.iso.org/standard/71183.html (accessed on 10 August 2026).
[6] Fang D, Brown N, De Wolf C, Mueller C. Reducing embodied carbon in structural systems: a review of early-stage design strategies. *J. Build. Eng.* 2023, 76:107054.
[7] Pomponi F, Moncaster A. Embodied carbon mitigation and reduction in the built environment—what does the evidence say? *J. Environ. Manage.* 2016, 181:687–700.
[8] Kirchherr J, Reike D, Hekkert M. Conceptualizing the circular economy: an analysis of 114 definitions. *Resour. Conserv. Recycl.* 2017, 127:221–232.
[9] Ellen MacArthur Foundation. Completing the picture: how the circular economy tackles climate change. 2019. Available: https://www.ellenmacarthurfoundation.org/completing-the-picture (accessed on 10 August 2026).
[10] Di Maio F, Rem PC. A robust indicator for promoting circular economy through recycling. *J. Environ. Prot.* 2015, 6(10):1095–1104.
[11] Khadim N, Agliata R, Marino A, Thaheem MJ, Mollo L. Critical review of nano and micro-level building circularity indicators and frameworks. *J. Clean. Prod.* 2022, 357:131859.
[12] buildingSMART International. Industry Foundation Classes (IFC). 2026. Available: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/ (accessed on 10 August 2026).

[13] Eastman C, Teicholz P, Sacks R, Liston K. *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*, 3rd ed. Hoboken: John Wiley & Sons, 2018.

[14] Pauwels P, Terkaj W. EXPRESS to OWL for construction industry: towards a recommendable and usable ifcOWL ontology. *Autom. Constr.* 2016, 63:100–133.

[15] Zhang C, Beetz J, de Vries B. bimSPARQL: domain-specific functional SPARQL extensions for querying RDF building data. *Semant. Web* 2018, 9(6):829–855.

[16] buildingSMART International. Information Delivery Specification (IDS). 2026. Available: https://www.buildingsmart.org/standards/bsi-standards/information-delivery-specification-ids/ (accessed on 10 August 2026).

[17] buildingSMART Technical. Referencing bSDD in IDS and IFC. 2026. Available: https://technical.buildingsmart.org/services/bsdd/referencing-bsdd-in-ids-and-ifc/ (accessed on 10 August 2026).

[18] European Committee for Standardization (CEN). EN 17412-1:2020. Building information modelling—level of information need—part 1: concepts and principles. 2020. Available: https://www.nen.nl/en/nen-en-17412-1-2020-en-277555 (accessed on 10 August 2026).

[19] Xue F, Wu L, Lu W. Semantic enrichment of building and city information models: a ten-year review. *Adv. Eng. Inform.* 2021, 47:101245.

[20] Meex E, Hollberg A, Knapen E, Hildebrand L, Verbeeck G. Requirements for applying LCA-based environmental impact assessment tools in the early stages of building design. *Build. Environ.* 2018, 133:228–236.

[21] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 2002, 6(2):182–197.

[22] Xue Q, Wang Z, Chen Q. Multi-objective optimization of building design for life cycle cost and CO_2 emissions: a case study of a low-energy residential building in a severe cold climate. *Build. Simul.* 2022, 15(1):83–98.

[23] Ascione F, Bianco N, De Stasio C, Mauro GM, Vanoli GP. A new comprehensive approach for cost-optimal building design integrated with the multi-objective model predictive control of HVAC systems. *Sustain. Cities Soc.* 2017, 31:136–150.

[24] Kamazani MA, Dixit MK. Multi-objective optimization of embodied and operational energy and carbon emission of a building envelope. *J. Clean. Prod.* 2023, 428:139510.

[25] Jalaei F, Jrade A, Nassiri M. Integrating decision support system (DSS) and building information modeling (BIM) to optimize the selection of sustainable building components. *J. Inf. Technol. Constr.* 2015, 20:399–420.

[26] Madireddy S, Gao L, Ud Din Z, Kim K, Senouci A, *et al*. Large language model-driven code compliance checking in building information modeling. *Electronics* 2025, 14(11):2146.

[27] Saka AB, Oyedele LO, Akanbi LA, Ganiyu SA, Chan DWM, *et al*. Conversational artificial intelligence in the AEC industry: a review of present status, challenges and opportunities. *Adv. Eng. Inform.* 2023, 55:101869.