CASING THE JOINT

Predicting Embodied Carbon in Early-Stage Geometries through Case-Based Reasoning

Z. HEFFERNAN,
UNSW Built Environment, Sydney, Australia
z.heffernan@unsw.edu.au

Keywords. Predictive methods, regressive modelling, embodied carbon, early design stages, Case-Based Reasoning (CBR).

Abstract. Embodied carbon (EC) is difficult to quantify in the early stages of design without detailed design information. In the later design stages, when EC analysis is traditionally performed, higher sunk costs can be incurred, and key design decisions are often irreversible. Additionally, both clients and designers are becoming increasingly aware of the environmental impacts of their projects and how they can meet or exceed sustainable energy/environmental performance targets. A process that assists with predicting EC based only on simple early models could help to ground sustainable design principles in all early stage decision-making as well as influence the likelihood of realising more sustainable design outcomes overall. The tool this research proposes would utilize a subset of machine learning called case-based reasoning (CBR), in which a database of previous examples is searched using regressive modelling for applicability and similarity to an early-stage geometry, and known EC values are applied to the new geometry parametrically to give an estimate of its predicted EC value. The following research builds on the recent uptick of applying machine-learning (ML) methods to the architecture and construction industry, and extends it further by linking CBR methodology to supporting these sustainable design decisions.

1. Introduction: Research Context and Motivation

It has been apparent for decades that human activities have drastically influenced the Earth's climate. The construction of new buildings accounts for up to 39% of energy-related carbon dioxide emissions, 36% of global energy use, and 24% of the raw materials extracted from the Earth's lithosphere, annually (Dean et. al. 2017, Bribian, 2020). These building and construction industry emissions figures are particularly worrying given the United Nations Framework Convention on Climate Change (UNFCCC) estimates that 230 billion square metres of new construction will be built in the next twenty years – the equivalent of the entire city of Paris *every week* (Birol, 2017). Existing approaches to addressing EC are valuable from a statistical standpoint, and in their current form they help designers understand the extent of EC in a building in the final stages of design. But these approaches are limited by their need for detailed project data, hence their prevalence in the later stages of design as reporting tools, as opposed to design-support tools. EC tools are well positioned to better assist designers appreciate the impact of EC content in building materials both more generally and in relation to the specific geometries they propose, and help influence the likelihood of lower EC content in building projects overall. There is, therefore, incentive to actively research new methods of EC estimation which can contribute to reducing the carbon impact of new buildings and develop tools which enable designers to be aware of the effects their decisions have on this carbon impact, earlier and with more information at hand.

EC, specifically in materials, refers to the carbon expended and emitted in their production, construction and end of life phases, but not in their use phases (Augusti-Juan, I 2017). It is both in the designer's and stakeholder's best interests that the carbon cost of a building be reduced as much as possible during a project, for both financial and climate-conscious motivations. Currently, traditional methods of EC analysis, such as Tally for Revit, are typically performed in the later stages of design, as this represents a time in the process where the most accurate and complete data is available. EC analyses completed at this time have the benefit of being accurate and detailed but are limited in many ways.

> "The spreadsheet-based methodology was simply too slow to be part of the actual design process… by the time you were done, the design process was already over."
>
> (KieranTimberlake, 2014)

The opportunity to change design elements, should the EC analysis reveal unfavourable results, becomes increasingly constrained as the project nears completion. Carbon-reducing measures are either implemented as a depthless, stopgap approach or not at all. Measures that *are* implemented in these later stages of design take considerable time, effort, and cost to deploy. Poor or non-existent decision-support tools have influenced the struggle of the architecture, engineering, and construction (AEC) industry to reconcile the need to be aware of the impacts of embodied carbon while simultaneously having the ability to easily change design elements before they are finalized. In other words, designers need an integrated process that can keep up with the rapidly evolving decisions made in the early stages of design.

When developing an early-stage EC decision-support tool for use in the AEC industry, machine learning (ML) can potentially offer an innovative way to process and analyse large amounts of data autonomously, a traditional barrier for tool development in this area. ML has the unique ability to scale with the complexity and speed required for any given project, while existing EC tools tend to slow down significantly when faced with larger datasets (KeiranTimberlake, 2014) The computational design field has more recently being to explore a range of applications for ML in design technology tools, and the predictive capabilities ML possesses, while yet to be widely accepted and proven in practice, have the potential to be a valuable resource for designers (Khean. 2018) Using a database derived from previously completed projects, for example, there is the opportunity to predict the EC value for early-stage geometry that does not yet contain detailed design information, based on the geometry's spatial and proposed material similarities to existing designs. Such a tool would utilize Case-Based Reasoning (CBR), an umbrella term that describes the process by which humans - and increasingly, machine learning processes – draw from the outcomes of previous experiences to solve problems in the present (Richter, 2011).

The methodology and the subsequent CBR tool were developed in collaboration with industry partner Bates Smart, with the goal of helping designers and clients make more informed and environmentally conscious designs. Bates Smart participated in defining the initial research problem and provided feedback throughout the iterative development process.

2. Research Aims

The primary aim of this research project is to develop a parametric Case-Based Reasoning tool with the ability to draw from a database of existing floor plans and predict the EC value of a novel floor plan, with comparable accuracy to existing EC reporting tools.

The following research will aim to assess, a) the accuracy and feasibility of CBR for the purpose of predicting EC in early-stage geometry, b) the overall speed and performance of the tool, and c) the theoretical benefits of such a tool, by discussing the potential effectiveness of a workflow that affords designers and increased insight into the impacts of their material and construction choices in the early stages of design.

3. Research Question

The following research aims to address the reduction of carbon content in early stage design using a new method – Case-Based Reasoning:

How accurate and effective is CBR as a method to develop an Embodied Carbon prediction tool for use in early stage design for novel geometries?

4. Methodology

This research adopts the overarching methodology of Action Research (AR), while also addressing the growing problem of sustainable design in general. AR is an appropriate methodology in this case, as it encourages research to be undertaken with a pragmatic goal – ie: to address real-world issues and present realistic solutions. Bates Smart, the industry partner for this research project, helped to ground the research in a professional and realistic context by offering an insight into the feasibility, benefits, drawbacks, and relevance of this research and the subsequent tool in an industry setting. The methodology of participatory AR, as described by Baskerville drawing on Lewin, is characterized by the involvement of practitioners and industry professionals as “both subjects and co-researchers” (Baskerville 1999, Lewin 1946).

AR is described by Lewin as a “change-oriented approach … grounded in an acute observation of the effects of change being introduced into a system” (Lewin, 1946). It “simultaneously assists in practical problem solving and expands scientific knowledge” (Baskerville, 1999). In short, AR aims to enact change and generate knowledge about the change. Because of the proactive nature of AR methodology, an AR researcher must consider themselves part of the study and must account for their own biases, limitations, and motives (Azhar et. al. 2009), as well of those of their stakeholders and contributors. AR is also highly interpretive – assumptions are drawn from an observation of the fundamental change introduced by the observer, and because of this, as O’Brien notes, AR first requires a definition of scope or else it has the potential to stray too far (or not far enough) from the research itself and into ‘action’ (O’Brien 1998).

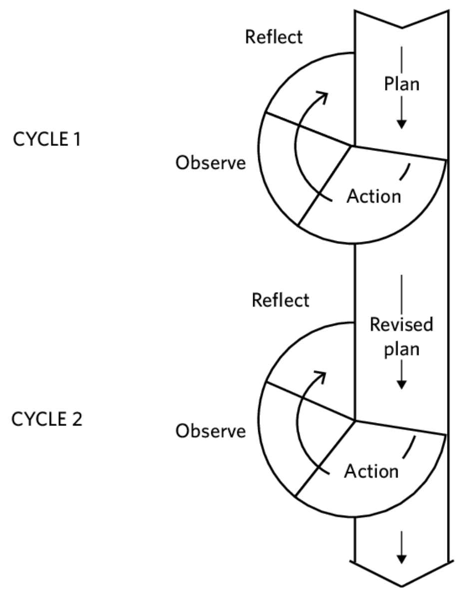


Figure 1. Action Research ‘spiral’. (Kemmis and McTaggart, 1988, p.44)

AR informs the approach to this research project in terms of its iterative development and aim of simultaneously expanding knowledge and developing pragmatic methods of EC analysis to address sustainable design objectives. The research project has been developed in collaboration with industry partner Bates Smart, who have contributed to identifying the research problem and provided feedback throughout the development of the research and CBR tool.

5. Background Research/Literature review

5.1. PREDICTIVE METHODS

While CBR is used extensively to solve problems in all aspects of human life, the computational design field has only recently begun to adopt it as a form of machine learning. The predictive capability it possesses, while yet to be widely accepted and proven in practice, has great potential to become a valuable ML approach for designers (Khean 2018) However, as Khean et. al. observes, "the inherent knowledge gap between the fields of architecture and computer science has meant the complexity of machine learning, and thus its potential value and applications in the design of the built environment remain little understood" (Khean 2018). While many subsets of ML such as genetic evolutional algorithms, mass data gathering and manipulation, and Bayesian neural networks, among others, have been applied to architectural practice before, there are limited examples that focus on developing a predictive model and methodology specifically for the early stages of design (Eisenstadt 2019). Predictive methods, however, have been applied to generative floor plan iterations (Eisenstadt 2019), operational carbon reduction (Victoria & Perera 2018 (1)), and parametric material choices (Loveridge 2011), providing precedent for a subsequent method to be developed. Using traits from these existing examples, an ML method can be developed to address the problem of embodied carbon prediction in novel geometries in the early stages of design.

Predictive models in mathematics are typically underpinned by a linear regression method, allowing users to visualize how multiple inputs are mapped to a desired output (Budig 2020). For instance, Budig's proposed method uses a regression approach to find a mathematical function that maps generic quantitative building parameters (such as Gross Floor Area, perimeter, material components, etc.) to different material volumes and spatial types, which are later used to visualize embodied carbon per material component in addition to the building's total GWP. In particular, material choice is not often considered alongside environmental impact until the later stages of design, when it is harder and costlier to change major design features (Bates 2013). As Bates also notes, there is a "need for environmental impact data in 'real time' … at the same pace at which they [designers] make decisions" – ie: not after the majority of a building's design has been finalized (Bates 2013).

5.2. MATERIAL TYPES, DATA COLLECTION, AND INPUTS

Material choice has increasingly become more prominent in the hierarchy of early stage design decisions as both clients and designers have become more aware of sustainability issues and the impact of common building materials on the environment (Loveridge 2011) . Buildings that are designed and constructed with sustainability in mind from the early stages of design have, on average, 40% less GWP than buildings designed and constructed using traditional methods (Bhochhibhoya 2017). Additionally, innovation in the field of architecture and engineering has produced increasingly more sustainable materials, as well as enabling the use of sustainable materials where previously it has not been feasible to do so – for example, the use of cross-laminated timber as a structural element in large towers (Kimpian 2009).

In relation to a potential predictive tool, the method chosen to collect and collate building material data as an inputs prior to predictive modelling has an impact on the reliability of the output, and can influence the users perception and awareness of a material's total contribution to the buildings embodied carbon value. It is therefore important that data collection methods remain as accessible, reproduceable and understandable to designers as possible. Budig et. al. emphasizes this point, and advocates for the use of Industry Foundation Classes (IFC) files as an "excellent source of information", as they contain quantitative data on material quantities and buildings shapes that can be transcribed into a workable database automatically (Budig 2020).
In order to further simplify the collection of material data for use in embodied carbon analysis, Victoria & Perera (2018(2)) further categorized building elements from 28 office buildings by ranking their embodied carbon value in order to find "carbon hotspots" – ie: elements whose embodied carbon value contributed the most to the building's total– and it was found that External

Walls, Services (lift cores, HVAC, electrical, plumbing, etc.), and Upper Floors contributed to the total embodied carbon value at a ratio of 80:36 (Victoria & Perera, 2018 (2)). Research of this type is fundamentally important when considered in conjunction with case-based reasoning and predictive models as it provides insight into the required specificity and format of potential ML inputs.

5.3. CASE-BASED REASONING AND EMBODIED CARBON

Case-based reasoning has the potential to be an effective ML method to predict values in new geometry when given inputs from existing buildings. The benefits of developing a tool to predict EC in the early stages of design include: reducing expenses over the course of a project by minimising late changes to detailed design drawings (Eisenstadt 2019), as well as providing an additional layer of data from which designers and clients can enact sustainable and informed material choices earlier in the design process (Budig 2020). Current research on the developing field of ML in computational design suggests that designers appreciate the outcomes of an ML-based tool but there exists a gap in understanding when it comes to the “inner workings” of such a tool (Khean 2018). In addition, designers require a tool that works “in real-time” with their material selection process in order to assist in the choice of sustainable materials when it is most feasible to compare different materials – in the early design stages (Bates 2013). Considering the findings of the studies mentioned in this review, there is both precedent and potential to develop such a tool.

6. Case Study

The following case study aims to resolve the stated research aims through the development of a parametric tool that utilizes CBR methods to predict the EC content of a novel floor plan geometry. The practical feasibility of this tool in its current form can be assessed via one key metric at this stage in the development process: accuracy. In this case, accuracy is measured comparatively to known EC values. The novel geometry used throughout the process has an embodied carbon value that is already known, but not added to the database, so as to provide a benchmark from which to ascertain the accuracy of the CBR process’s prediction.

The CBR tool was developed primarily in Grasshopper, a visual-scripting workspace for the CAD tool Rhino. Initial data manipulation and setup was conducted in Revit and Excel, while the post-CBR workflow was visualized in Adobe InDesign and Photoshop to present the benefits of a potential interface UI.

6.1. CREATING THE DATABASE

The initial process of developing a database was primarily derived from a combination of factors presented by Budig et al. (2020) and Eisenstadt (2019) in their research on similar topics. The creation of a database is a crucial first step as it plays a large role in defining the format and quantity of the inputs for the CBR script. BIM programs offer a valuable source of information from which to build the initial database, in the form of IFC files (Industry Foundation Classes), which can be broken down into the key data points relevant to this tool. In this case, Revit was used to produce the input data as it represents a popular industry choice for architects, designers, and engineers, as well as being the software of choice for industry partner Bates Smart.

6.1.1 Initial Parameters

The choice was made early in the iterative design process to limit the scope of the CBR tool, primarily to remain within the 10-week time frame, but also in consideration of potential overlap with previous research. As such, the tool will draw data from single floor plans that represent a typical ‘slice’ of commercial/office buildings. The goal of the tool is to draw a prediction from a data-sparse early-stage model, so the input parameters chosen were done so because they are most commonly available in early-stage building models, according to initial consultations with the industry partner.

The parameters chosen for inclusion in the database were:

1. Area (m^2)
2. Perimeter (m)
3. Material volume (m^3)
    a. Concrete
    b. Steel
    c. CLT Timber
4. Spatial Typology Areas (m^2)
    a. Working

    b. Circulation
    c. Services
5. Construction System (categorical.)
    a. Shear Wall only
    b. Column Beam only
    c. Shear Wall and Column Beam

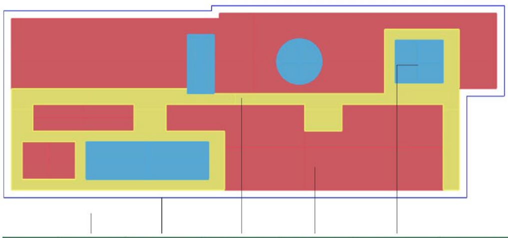
These parameters can be classified into categorical and quantitative parameters. While it is possible for CBR algorithms to account for categorical conditions by describing them as a mathematical vector (in a process known as “One-Hot Encoding” (Friedman 2001)) , it is not necessary for this tool as in this case it constitutes only one parameter – Construction Systems – which can be readily tracked alongside the CBR algorithm throughout the script. However, it is important to clarify the following quantitative parameters:

Spatial Typology Areas (Working, Circulation, and Services)

The spatial typology definition method was adapted from Budig et al. (2020), (in which the researchers used circulation/served/serving) as a way to identify and quantify the purpose of rooms/areas within a floor plan. It was important to visualize the different area-specific carbon embodiment for comparison in the final visualization, as well as for comparative analysis by the similarity algorithm, as described later. Working spaces are defined as being areas which are productive in some way – cubicles, offices, workshops, etc., while Circulation spaces are areas which link other areas together or, for the sake of simplicity and similarity of material components, serve no purely productive purpose (hallways, passages, lounges, break rooms), while Services constitute the areas vital to building structure, maintenance, or utility (liftcores, stairwells, bathrooms, and storage).

6.1.2 Data Extraction from Revit

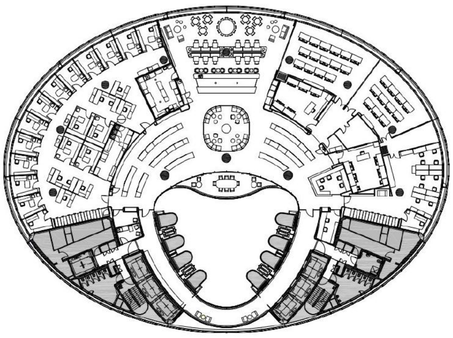
For a preliminary study, the data was manually extracted and collated in Revit into the following classes:



Case	Area (m2)	Perimeter (m)	SA Circulation	SA Working	SA Services	Concrete (m^3)	Steel (m^3)	CLT (m^3)	Other (m^3)	Total Material Volume (m^3)
Floor1	1600	160	0.27	0.54	0.19	13.5	11.65	18.25	1.45	44.85

Figure 2. Diagrammatic representation of the key parameters extracted from Revit.

Bates Smart presented a floor plan from their own project history, with embodied carbon data attached (Figure 3).



Environmental Impact Totals	Product Stage [A1-A3]	Construction Stage [A4]	Use Stage [B2-B5]	End of Life Stage [C2-C4]
Global Warming (kg CO_2eq)	-8,698	154.3	0	7,007
Acidification (kg SO_2eq)	22.35	0.7152	0	26.94
Eutrophication (kg Neq)	1.311	0.05823	0	7.007
Smog Formation (kg O_3eq)	296.6	23.63	0	90.60
Ozone Depletion (kg CFC-11eq)	3.745E-004	5.286E-012	0	5.623E-011
Primary Energy (MJ)	109,932	2,244	0	3,829
Non-renewable Energy (MJ)	38,053	2,191	0	3,600
Renewable Energy (MJ)	71,878	54.27	0	227.7

Figure 3. The office floor plan provided by Bates Smart, with the relevant data points shown in a table.

The process of extracting these datapoints from Revit was initially planned to be automated through Tally, an existing LCA tool that 'tallies' material volumes as well as the basic building parameters needed for the script. These parameters could then be exported in Excel format. However, the prohibitive cost of a Tally license lead to the data needing to be transcribed directly from Revit to Excel manually (Figure 4). This process was tested first on the Bates Smart floor plan, before being refined and applied to the remaining floor plans, as discussed below.

The target database size was 15 floor plans, to account for variation and provide a large enough dataset from which accurate predictions could be drawn. Budig et al. (2020) and Eisenstadt (2019) both used dataset with 10-

20 cases for preliminary studies, so the target number of 15 was appropriate for this similar research.

	A	B	C	D	E	F	G	H	I	J	K
1	Case	Area (m2)	Perimeter (m)	SA Circulation	SA Working	SA Services	Concrete (m^3)	Steel (m^3)	CLT (m^3)	Other (m^3)	Total Material Volume (m^3)
2	Floor1	1600	160	0.27	0.54	0.19	13.5	11.65	18.25	1.45	44.85
3	Floor2	950	120	0.21	0.65	0.14	27.55	5.25	4.65	2.2	39.65
4	Floor3	500	220	0.23	0.48	0.29	9.5	18.2	3.5	2.1	33.3
5	Floor4	1762	234	0.27	0.54	0.19	13.5	11.65	18.25	1.45	44.85
6	Floor5	234	567	0.48	0.65	0.21	27.55	5.25	4.65	2.2	39.65
7	Floor6	567	220	0.54	0.48	0.23	9.5	18.2	3.5	2.1	27.55
8	Floor7	1532	1612	0.65	0.54	0.27	13.5	11.65	18.25	1.45	9.5
9	Floor8	850	850	0.48	0.65	0.14	27.55	5.25	4.65	2.2	13.5
10	Floor9	475	475	0.23	0.48	0.29	9.5	18.2	3.5	2.1	39.65
11	Floor10	1674	160	0.27	0.54	0.19	13.5	11.65	18.25	1.45	33.3
12	Floor11	955	120	0.21	0.65	0.14	27.55	5.25	4.65	2.2	27.55
13	Floor12	542	220	0.23	0.48	0.29	9.5	18.2	3.5	2.1	9.5
14	Floor13	1612	160	0.27	0.54	0.19	13.5	11.65	18.25	1.45	13.5
15	Floor14	234	120	0.21	0.65	0.14	27.55	5.25	4.65	2.2	39.65
16	Floor15	500	220	0.23	0.48	0.29	9.5	18.2	3.5	2.1	33.3

Figure 4. Excel spreadsheet with transcribed data from Revit.

6.2. DEFINING SPATIAL TYPOLOGIES

To achieve a workflow in which geometry can be directly referenced in the subsequent Grasshopper script, the floor plans (which will henceforth be called 'cases', as the term will be used when discussing the algorithm in the following section) needed to be imported into Rhino, the CAD software which Grasshopper runs alongside. This process was completed by hand, but in future iterations could be completed using image-recognition AT (although this is outside the scope of this research).

The cases were then assigned into an appropriate layering hierarchy, to simplify the referencing process in Grasshopper.

Figure 5. The 15 cases scaled to the correct size in Rhino.

The cases, once transcribed into Revit, were traced once more to define the boundaries for their unique spatial typologies. Working, Circulation, and Service areas were traced by hand to define these boundaries, and then assigned as a sublayer for their parent case, as described below:

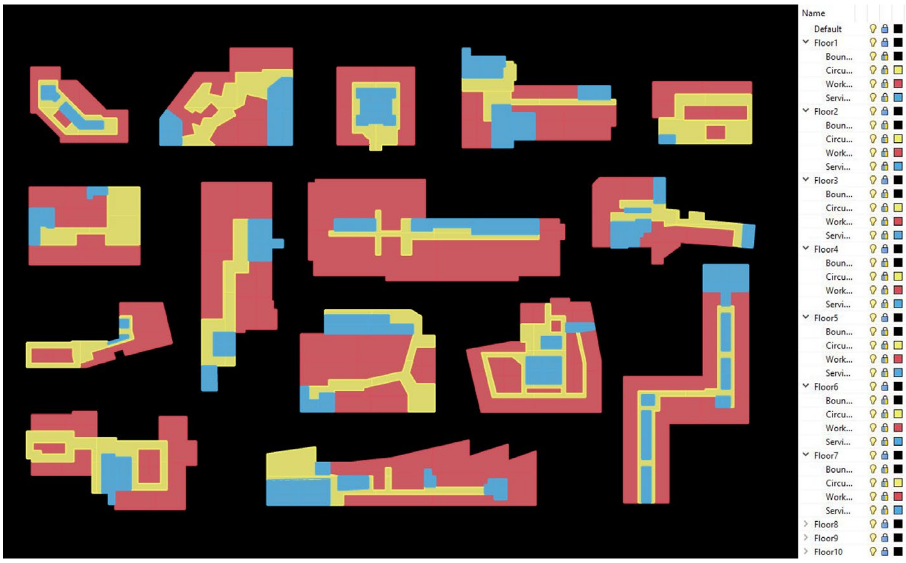


Figure 6. Spatial Typologies defined and assigned into parent/child layers.

6.3. DEFINING INPUT PARAMETERS

6.3.1 Referencing geometry and parameters

At this point in the process, the ‘case’ database had been set up in two formats: geometry and parameters, in Rhino and Excel, respectively. A method was needed to collate and combine geometric and numerical data from both sources for later use by the CBR algorithm (Figure 7). A Grasshopper plugin, called Bumblebee, was used to import and decode the Excel data into a readable format for the CBR script. The Rhino geometry was ‘live-referenced’ using a plugin called Human and a component called Dynamic Geometry Pipeline, which extracts geometric data from the predefined Rhino layers discussed in the previous section.

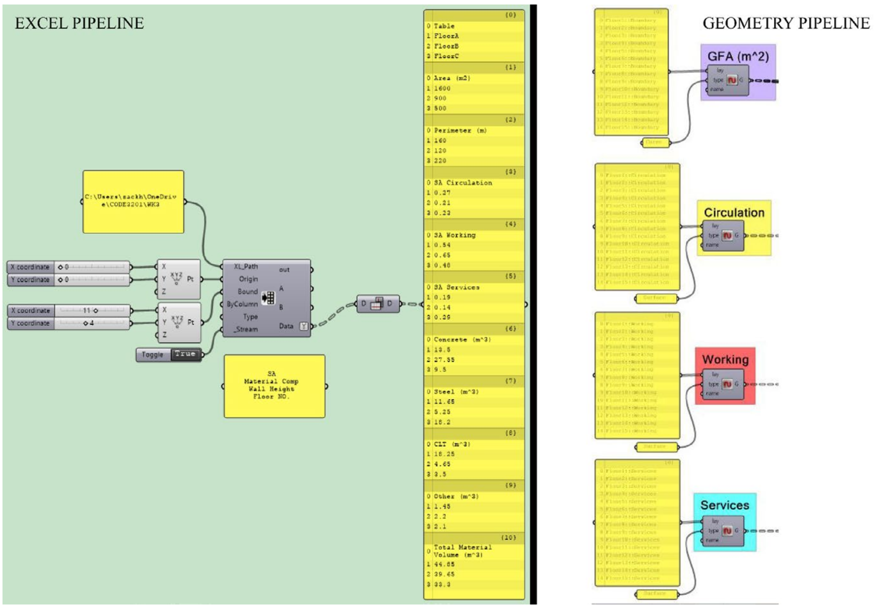


Figure 7. Both the geometry and parameters were referenced for use in the Grasshopper-based CBR script.

6.3.2 Defining the CBR inputs

Case-Based Reasoning is unique in that its complexity can be defined and appropriately scaled before writing the algorithm (Poole 2018). If the 'target' value is a simple numerical prediction, and the number of 'cases' is small, the most appropriate regression method is the '*k*-nearest neighbors' method, for some variable '*k'* (Poole 2018). It is important to note that this method is only applicable when all the input cases have a *'k'* value in the same format as the anticipated '*k'* outcome.

The target case geometry was also defined in this stage of the process. One of the Bates Smart floor plans was used, as the EC content of it was precisely known. This value was then removed from the dataset, but noted externally, for reference in the final stage to check the accuracy of the CBR method.

6.3.2 Global Warming Potential as an output of the CBR Method

In this case, the anticipated input and outcome are the EC content, measured in $KgCO_2$ *eq per* m^2 (kilograms of carbon dioxide equivalent per metre squared).

CO_2 eq is a metric used for describing the impact of a ‘bundle’ of materials in a single unit. It signifies that 1m^2 of CO_2 that would have the same GWP (Global Warming Potential) as *x* m^2 of the ‘bundle’ of materials. For example, CO_2 has a GWP of 1, while structural concrete has an average GWP of 30.9 (Bhochhibhoya, 2017). Heavily simplified, but appropriate for the purposes of comparison, this means that structural concrete has 30.9x more GWP than the same amount of CO_2 (usually measured with a duration of 100 years).

Table 1. Comparison of KgCO2 eq per m2 of the construction types and materials used in this CBR script, data from Hammond et al (2011), and Budig et al. (2020) .

Material	**Concrete (UK avg.)**	**Steel (UK avg.)**	**Timber (UK avg.)**
$KgCO_2eq$ per kg^2	860.0	146.0	41.0
Construction System	**Column/Beam**	**Shear Wall**	**Column/Beam and Shear Wall**
$KgCO_2eq$ per m^2	12.96	9.17	30.90

6.4. CASE-BASED REASONING ALGORITHM

6.4.1 CBR Method - Retrieve

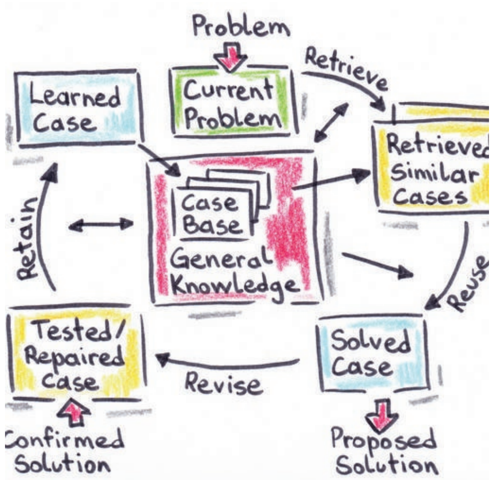


Figure 8. Diagrammatic representation of the CBR algorithm.

Reduced to a fundamental workflow, CBR follows the following '4 R's cycle':

Retrieve: Given a new target case, retrieve similar cases from the dataset.
Reuse: Adapt the retrieved case to fit the target.
Revise: Evaluate the solution and revise it based on how well it works.
Retain: Decide whether to retain this new case in the dataset.

The final 'R' – Retain – is not applicable for the scope of this preliminary research but would be useful in an integrated workflow in which 'solved' cases are recycled for future use.

The CBR method depends on a target parameter (X_i) being assigned a 'weight' (w_i), and a distance metric being applied to measure the 'closeness' of each set of two examples (the 'target' and each 'case', consecutively). The closeness according to the weighted value of the target parameter can be measured using a variation of the Euclidean Distance formula (1).

$$d\,(e_1, e_2) = \sqrt{\textstyle\sum_i \; w \, * \left(Xi(e1) - Xi\,(e2)\right)\text{^}2} \tag{1}$$

The formula can be applied directly into the equation component in Grasshopper, and the input variables can be defined and linked to the previous Excel output. Alternatively, grasshopper has a native component that replicates the Euclidean Distance formula, called the Similarity component (Figure 9). Two cases and a weighting value were the inputs, while the Euclidean distance, or the 'similarity' was the output. This was then used to check the similarity of each case, and by process of elimination, find the most similar case to the target.

The target parameter in this case was the ratio of GFA (m2) to each Spatial Typology (m2), ensuring that the chosen case has the highest likelihood of possessing similar material composition to the target, without knowing exactly what those compositions are, due to the nature of working with early stage geometry. This essentially comprises the first 'R' of the cycle – Retrieve.

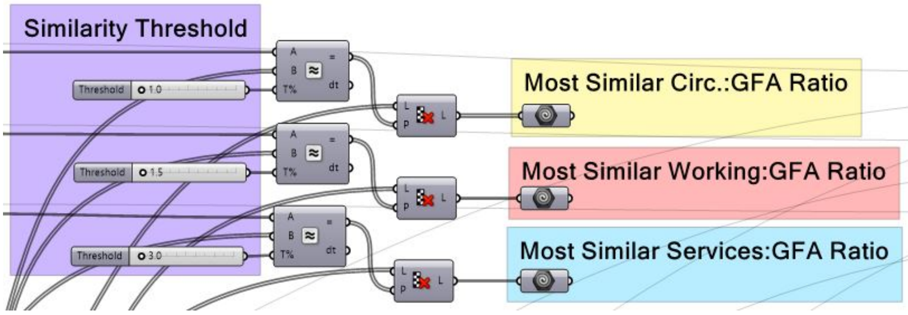


Figure 9. The 'Similarity' Component, which isolates the most similar case to the target, and a comparison of a weighting of 1 vs 1.5 vs 3.

6.4.2 CBR Method – Reuse

The next step in the CBR method was to adapt the similarity-tested retrieved case to the parameters of the target case. Each parameter was multiplied by a similarity ratio ('case':'target') in order to estimate the material volumes and subsequent EC content of the chosen case.

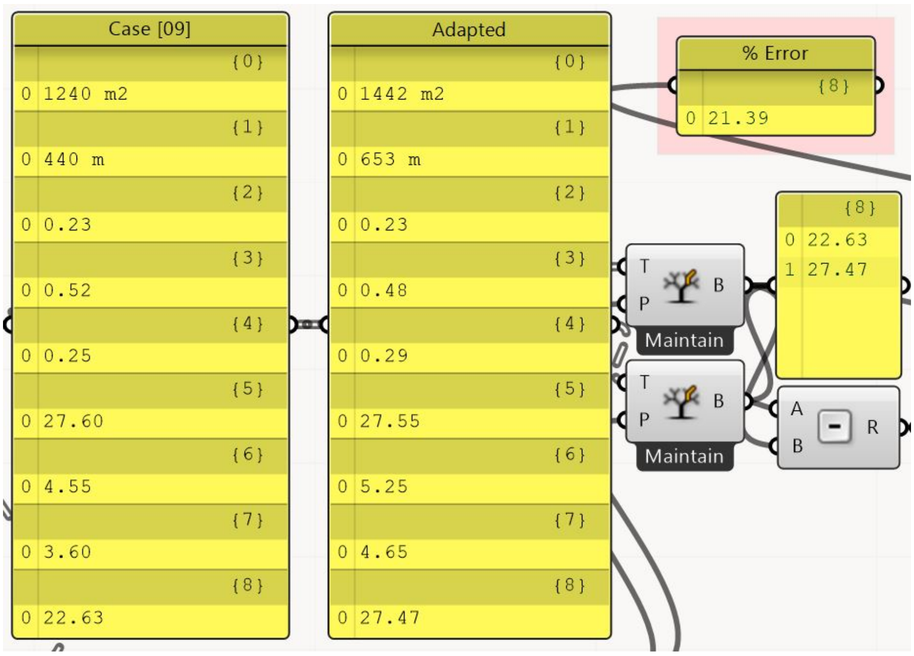


Figure 10. Adapted parameters for the target case.

6.4.3 CBR Method – Revise

The approximate accuracy of the CBR method was measured by comparing the output of the script to the known EC value of the target case (Figure 10.)

In summary, the CBR process used a database of 15 cases to evaluate the most similar geometry for a target case, based of a variety of early-stage parameters – area, perimeter, basic material composition, and spatial typologies – and returned a result with an error percentage of ~21%.

To determine the definitive accuracy and precision, the same process would need to be repeated on a wider range of input floor plans, which was not feasible given the limited amount of floor plans available during the time-frame of the research.

7. Discussion

The research has so far developed a comparison algorithm and a spatial area type visualization which assists with early stage design decision-making, specifically when comparing the impact of embodied carbon in construction types and materials. Drawing from a small database of existing office and commercial floorplans, the research and subsequent algorithm has successfully proven that Case-Based Reasoning is an appropriate application of machine learning for gaining autonomous and human-readable insights from databases of simple floor plan geometry. The algorithm developed throughout the 10-week time frame reached the stated aim of sorting a database of 'case' geometry by compatibility to a undefined input geometry - the 'target' - in order to predict, or more accurately, estimate, the embodied carbon content of the target based on generic design parameters similar to both. This success has potential positive implications for future work on this topic, and the groundwork has been laid out for integration into a real-life workflow.

While the framework, workflow and overall theory behind the prediction tool proved to be sound, the tool itself in its current form lacks the refinement needed to integrate it into the workflow of a contemporary designer. In the initial weeks of research discussion, the tool was proposed as a 'back-end' for a UI interface which would allow non-designers to quickly compare construction and material iterations for a project and clearly see which would be the most ideal to implement from an embodied carbon standpoint. In reality, this approach was tabled once the complexity of developing a CBR script from scratch was apparent. Instead, the research focus was shifted to exploring the theory and potential utility of CBR as a predictive tool, with a script being developed as a proof-of-concept to compliment the research. Cost and time were the two largest limiting factors, both of which inhibited the ability to fully implement machine learning into the script. A license for the Tally plugin for Revit proved to be outside of the scope and budget for this research, and its application in the translation of material data directly from Revit would have, by the nature of its inherent integration into the popular CAD software, been useful in the construction of the case database. In addition, while CBR proved to be an appropriate method for deriving comparative similarity from an external database, a longer timeframe for research would have allowed for the testing and comparison of more machine learning methods. In particular, a Bayesian Neural Network (BNN) was applied to a similar embodied carbon prediction task by Budig, et al. (2020), and as such a comparison between the effectiveness and accuracy of the two methods would have been valuable as a benchmark.

Contemporary research has been, to some extent, successful in showcasing the ability for machine learning processes to have a significant impact on the decisions of designers in the early stages of design, and this research paper endeavored to build upon this further. The key difference between existing research and the research undertaken in this paper is the level of complexity needed to achieve, in theory, similar results. The titular BNN used in Budig et al.'s research suffered from over-complexity when 'combing' through the initial database of cases, which was time and resource intensive (Budig et al. 2020). Case-Based Reasoning, being explicitly goal-oriented as opposed to probability-oriented, is more suited to smaller, well-defined datasets as it has the advantage of being set up to find explicit input parameters and where in the dataset to find them. Of course, as the dataset in this research was small in sample size, CBR has a clearer advantage. For larger datasets, which are unavoidable should the process be applied to the AEC industry as a whole in the future, the efficacy of CBR approaches will likely see diminishing returns.

While this research has elaborated and expanded upon existing early-stage prediction methods, it falls short of developing a human-accessible tool that could be put to use in a real-life workflow. With either more time and resources, creating a 'front-end' for the CBR script would be both realistic and feasible. Such an interface would not only reformat the EC results in a way that is accessible to anyone, but could also interact with the script as it runs in a background process, allowing users to change inputs in order to rapidly visualize the impacts of different construction systems or material choices in real-time. The CBR algorithm itself could also be updated to search for different goal parameters (ie: cost, or a specific material), or even a combination of parameters (ie: searching for the most similar design that both reduces embodied carbon but doesn't require a construction system to be changed, or a design that reduces embodied carbon but doesn't impact the overall volume of materials, for example).

8. Conclusion

The CBR method presents a novel and potentially viable method for estimating EC content in early-stage geometry without detailed design information available. Contemporary research has shown that designers and stakeholders studying LCA methods of EC estimation require a parametric tool which can rapidly estimate EC content from the early stages of the design process and is able to quickly update and work alongside the designer’s workflow, rather than impeding it. By using a simplified linear regression workflow, the process was able to quickly provide an estimated EC content for a test geometry to within ~22% of its actual value. While this particular process is not yet accurate or expansive enough for integration into real-life workflows, this study has shown that there is significant potential for further investigation to be conducted on the integration of ML methods and tools, to be used in parallel with conventional CAD software. The study also addresses the trending goal of designers to building sustainably and develop sustainable knowledge and forethought, a key issue that is becoming increasingly prevalent in modern practice.

The research explored in this paper lays the theoretical and practical groundwork for future developments that have the potential to fundamentally reshape the process of sustainable design thinking, and help both designers and stakeholders be aware of informed and sustainable choices before they become irreversible.

Acknowledgements

Thank you to Bates Smart Sydney for a weekly review of the work conducted in this study, with special mention to Stephen Green and Aaron Coats for their advice and contributions. Thank you also to Nicole Gardner and Daniel Yu for thematic and technical feedback which proved invaluable while writing this research.

References

Agusti-Juan, I, Hollberg, A, and Habert, G, 2017, ‘Integration of Environmental Criteria in Early Stages of Digital Fabrication’, *eCAADe 2017: Proceedings of the 35th International Conference of eCAADe,* Sapienza University, Rome, vol. 2, pp . 185-192.

Bates, R. Carlisle, S. Faircloth, B. and Welch, R. 2013, ‘Quantifying the Embodied Environmental Impact of Building Materials During Design: A BIM-based Methodology’, *PLEA2013 – 29th Conference,* Sustainable Architecture for A Renewable Future*,* Munich, Germany, pp. 1-5.

Bhochhibhoya, S, et. al. 2017, ‘The Global Warming Potential of Building Materials: An Application of Lice Cycle Analysis in Nepal’, *Mountain Research and Development vol.37,* International Mountain Society Journal (IMS)*,* Kathmandu University, Nepal.

Birol, F. 2017 ‘Preface: Towards a zero-emission, efficient, and resilient buildings and construction sector’, *Global Status Report,* Global Alliance for Buildings and Construction (GABC), UNFCCC, Paris, France.

Bribian, I.Z., Capilla, A.V., and Uson, A.A., 2011, ‘Life Cycle Assessment of Building Materials: Comparative Analysis of Energy and Environmental Impacts and Evaluation of the Eco-Efficiency Improvement Potential’, *Building and Environment, CIRCE 2010,* University of Zaragoza, Spain, pp. 1134-1140.

Budig, M, et. al. 2020, ‘Data Driven Embodied Carbon Evaluation of Early Building Design Iterations’, *RE: Anthropocene, Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2020,* CAADRIA, Hong Kong, vol. 2, pp. 303-312

Dean, B., et. al. 2017 ‘Toward zero-emission efficient and resilient buildings’, *Global Status Report,* Global Alliance for Buildings and Construction (GABC), UNFCCC, Paris, France.

Eisenstadt, V et. al. 2019, ‘Generation of Floor Plan Variations with Convolutional Neural Networks and Case-Based Reasoning’, *eCAADe SIGraDi 2019: Architecture in the Age of the 4th Industrial Revolution,* eCAADe, SIGraDi, and University of Porto, Portugal, vol. 2, pp. 79-84.

Hammond, G, et al. 2011, ‘Embodied Carbon’, *The Inventory of Carbon and Energy (ICE),* University of Bath in association with Carbon Trust UK and Energy Research Council UK, Bath, UK, pp.10-14

Khean, N et. al. 2018, ‘The Introspection of Deep Neural Networks – Towards Illuminating the Black Box: Training Architects in Machine Learning via Grasshopper Definitions’, *Learning, Adapting, and Prototyping; Proceedings of the 23rd International Conference of the Association for CAADRIA,* CAADRIA, Hong Kong, vol. 2, pp. 237-246.

KieranTimberlake PA, 2014, ‘Tally Revit Application Webinar’, available at: https://vimeo.com/84681066 [accessed 25/10/2020]

Kimpian, J. et. al. 2009, ‘Sustainably Tall: Investment, Energy, Life-Cycle’*, ACADIA 2009, reForm(): Building a Better Tomorrow,* ACADIA, Chicago, Illinois, USA, pp. 130-143.

Kemmis, S., McTaggart, R. 1988. ‘The Action Research Planner’. *Deakin University Press*, Melbourne. pp. 44, 5–6.

Kemmis, S. 2009. ‘Action Research as a practice-based practice’, *Educational Action Research,* Charles Sturt University, Bathurst, pp.463-474

Lewin, K. 1946, ‘Action Research and Minority Problems’*, Journal of Social Issues*, vol. 2, no. 4, pp. 34–46

Loveridge, R. 2011, ‘Parametric Materiality: Material Properties as a Catalyst for Design’, *Circuit Bending, Breaking and Mending, Proceedings of the 16th International Conference on CAADRIA,* CAADRIA, Hong Kong, pp. 165-174.

Poole, D., Mackworth, A. 2018, ‘Artificial Intelligence 2E: Foundations of Computational Agents’*, Cambridge University Press*, UK, vol 2, section 7.7

Victoria, M.F. and Perera, S. 2018, ‘Managing Embodied Carbon in Buildings: A Pareto Approach’, *Built Environment Project and Asset Management,* Robert Gordon University Aberdeen, UK, pp. 504-515.

Victoria, M.F. and Perera, S. 2018, ‘Parametric Embodied Carbon Prediction Model for Early Stage Estimating’, *Energy and Buildings,* Robert Gordon University Aberdeen, UK, vol. 168, pp. 106-119.