

DATA-DRIVEN EMBODIED CARBON EVALUATION OF EARLY BUILDING DESIGN ITERATIONS

MICHAEL BUDIG¹, OLIVER HECKMANN²,
AMANDA NG QI BOON³, MARKUS HUDERT⁴,
CLEMENT LORK⁵ and LYNETTE CHEAH⁶
^{1,2,3,5,6}*Singapore University of Technology and Design*
^{1,2,3,5,6}{michael_budig|oliver_heckmann|amanda_ng|clement_lork|
lynette_cheah}@sutd.edu.sg
⁴*Aarhus University*
⁴mhu@eng.au.dk

Abstract. In the early design phases, Life Cycle Assessment can assist project stakeholders in making informed decisions on choosing structural systems and materials with an awareness of environmental sustainability through their embodied carbon content; yet embodied carbon is difficult to quantify without detailed design information in the early design stages. In response, this paper proposes a novel data-driven tool, prior to the definition of floor plan layouts to perform embodied carbon evaluation of existing building designs based on a Bayesian Neural Network (BNN) regression. The BNN is built from data drawn from existing floor plans of residential buildings, and predicts material volume and embodied carbon from generic design parameters typical in the early design stage. Users will be able to interact with the tool in Grasshopper or as an online resource, input generic design parameters, and obtain comparative visualizations based on the choice of a construction system and its environmental sustainability in a 'shoebox' interface - a simplified three-dimensional representation of a building's primary spatial units generated with the tool.

Keywords. Regression; Bayesian Neural Network; High-Rise Residential Buildings.

1. Introduction

Human activities have grossly influenced the Earth's climatic conditions. The construction and operations of buildings can account for up to 36 percent of global energy use and 39 percent of energy-related carbon dioxide emissions annually (Dean et al. 2017). Together with rapid urban growth driving up demand for residential apartments, improving building design for maximised environmental sustainability has become more important than ever.

To support this endeavour, Life Cycle Assessment (LCA) is typically adopted as a methodology to evaluate the impact of embodied and operational carbon in buildings (ISO 2006). For which, the impact of embodied carbon is typically

quantified in a metric known as Global Warming Potential (GWP) - a measure of the greenhouse gas (GHG) emissions for the manufacturing of building materials. The process of LCA is usually applied to late stages of design, where the building information is well advanced and there is little possibility to make improvements due to larger sunk costs incurred. Quantifying the impact of embodied carbon (GWP) within a building in the early stages would allow designers to plan for and minimize the subsequent environmental impact of building designs. However, GWP is difficult to quantify in the early design stages due to uncertain information regarding the design project and the resulting material volumes required.

This paper proposes the development of an LCA-based tool to address the challenge of quantifying GWP without a floor plan design in the early LCA stage. More specifically, we focus at the stage 2 - Concept Design of RIBA Plan of Work 2013 (Sinclair 2013), where the material and construction systems are not yet defined. The approach presented builds up further on previous attempts (Meex et al 2018).

As a case study, construction systems such as hybrid concrete-timber systems are explored in this research to assess the potential reduction in GWP due to the supplementation of carbon-intensive concrete with a more sustainable material like timber. A predictive model that maps generic design parameters to eventual LCA results were built using data from existing building projects in a Bayesian Neural Network regression framework. Our proposed model implicitly relates the basic building dimensions, to kinds of spaces (circulation/served/serving), inner/outer enclosures, ratio of support and infill components, to the eventual probabilistic material volumes and embodied carbon estimations. In the resulting tool, the user will be able to input initial client-driven parameters, visualize and evaluate the LCA results based on different construction systems and choices of infill systems.

2. Background

LCA provides a quantitative assessment of the upstream and downstream burdens throughout the life cycle of the building. There are three levels of LCA with varying levels of detail to cater for different stages of design: screening LCA, simplified LCA and complete LCA (EeBGuide 2014). In practice, LCA is typically conducted near the completion of the building project, where changes are more difficult to be made and at a high cost (Treloar et al. 2000). Following the logic of the MacLeamy curve (MacLeamy 2004), the ability to 'optimize' the resulting embodied carbon of the building and reduced the cost of changes is the highest in the earliest stages of design.

Only a few LCA tools for early stage evaluations are currently available. Screening or simplified LCA tools such as CAALA (Meex et al. 2018), aid in the initial phases of building design. VPL-based tools like Bombyx (Basic et al 2019) and Integrated Design Model (Otovic et al 2016) aim to provide design guidance based on simplified building information as input and intuitive carbon evaluation as output. These tools, as opposed to complete LCA tools, focus on few prime building elements, user-specific life cycle stages, and cover a reduced number of impact categories. However, such tools do still require an existing design to obtain

LCA results, which is why they are only suitable for planners who already possess a certain level of expertise. Our proposed framework allows users to evaluate and formulate immediate conclusions to influence design decisions with only generic, quantitative parameters. Furthermore, the tool builds upon the common graphical representation of GWP data with the comparison of two or more design variants. Through highlighting the best performing and worst performing design variants, users will be aware of the scope for more sustainable design.

To overcome the lack of design information and expertise in the earliest design stages, predictive models are developed to assist users in quantifying embodied carbon (Victoria et al. 2018), operational carbon (Ferlito et al. 2015), and costing (Liu et al. 2013) throughout the lifespan of the building to improve environmental and cost efficiency. These predictive models are typically built on linear regression approaches that map a few inputs to the desired output, allowing users to explore the relationship between basic inputs and outputs without the need for an existing design (Victoria et al. 2018, Ferlito et al. 2015). However, such research has mostly been focused on predicting operational carbon (Gadezi et al. 2016) rather than embodied carbon (Victoria et al. 2018). For this paper we assume the following set of generic quantitative parameters that help to determine the total amount of all surfaces in a typical residential building: footprint area, depth, and width of the building, the number of units required in the building, building height and typical floor height, and structural/material choices.

The main contribution of this paper is a software tool that would be shared as an online resource to predict GWP evaluation for early design iterations of residential buildings. The software inputs the generic residential design parameters, and outputs a ‘shoebox’ representing the typical resulting architectural primary spatial units, and a breakdown of material quantities and GWP of the said representation. Figure 1 illustrates how the coupling of LCA with predictive modeling facilitates the ability to compare the impact of different quantitative design variants, as an alternative to the standard screening process where a floor plan design has to be generated before LCA can be conducted. This paper discusses the various steps to construct the predictive tool, which are data collection, data modeling, and visualization of the building ‘shoebox’. The following sections describe each of the components illustrated in figure 1.

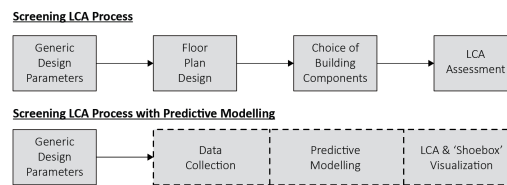


Figure 1. Proposed improved simplified LCA process with predictive modelling .

3. Data Collection

Data from numerous built projects with information on building parameters, material volume and GWP information of materials would be required to build a reliable predictive model. Industry Foundation Classes (IFC) files that commonly used in Building Information Modeling (BIM) programs are an excellent source of information. Such files include information on spatial elements, material quantities, and building shapes that could be automatically transcribed into a digital database. For a preliminary study, we manually collected information from case studies of built construction as illustrated in Floor Plan Manual Housing (Heckmann and Schneider 2018) and Manual of Multi-Storey Timber Construction (Kaufmann et al. 2018). State-of-the-art timber construction systems were compiled on top of other conventional systems for multi-storey apartment buildings as timber has substantial impact on lowering GWP. 17 case studies with two forms of construction system - Shear Wall and Column Beam systems and the concomitant material combinations implemented in these systems were studied. The data extraction process is shown in figure 2. In this process basic building dimensions, spatial areas (circulation/served/serving), inner/outer enclosures, ratio of support and infill components are measured and consolidated to prepare for the modelling process. Specifically, superstructure, structural (support systems) and non-load bearing elements (infill systems) were considered, as shown in table 1. The choices for both infill types provided to the users were adopted from Quartz dataset (Quartz,). The thicknesses of the respective structural walls were assumed to be consistent throughout the building.



Figure 2. Data extraction process: selection of case studies, translation and data entry. .

Due to the uncertainty of the service lifetime and end-of-life process, the scope of the LCA study is kept with the cradle-to-gate boundary, denoting the embodied carbon captured in raw material extraction, transportation, and manufacturing process of the building products. This boundary does not include the impact of transportation to the construction site. The LCA database is sourced from Quartz and GWP is adopted as an indicator for the measurement of the embodied carbon factor (Quartz). The functional unit will be $KgCO_2eq \text{ per } m^2$ with m^2 representing the gross floor area of the building.

Table 1. Building elements and respective materials accounted for embodied carbon in the case studies .

Type of System	Type of Building Component	Material
Support Systems - Shear Wall System (Superstructure structural)	Vertical load-bearing element 1	Concrete
	Vertical load-bearing element 2	Concrete
	Horizontal structure (Slab or frame)	Concrete, CLT, Glulam
Support Systems - Column Beam System (Superstructure structural)	Vertical load-bearing element 1	Concrete
	Vertical load-bearing element 2	Glulam, Steel
	Horizontal structure (Slab or frame)	Concrete, CLT, Glulam
Infill Systems (Superstructure non-structural)	Partitions	Drywall, Timber frame wall, Wet wall
	Frames	Panel, Glazing, Door, Window

4. Modelling Process

The goal of the modeling process is to predict the material volume for different inputs, before calculating the total GWP. A typical approach to prediction is a regression approach, which involves finding a mathematical function to map inputs to outputs (Friedman et. al, 2001). In our process, the inputs refer to generic quantitative building parameters (figure 3), and the outputs refer to the different material volumes and building areas mentioned in figure 4. Input data can be categorized into unit level and building level information:

1. Unit level inputs influence the size and programmatic constraints of one apartment unit in the building: Number of served spaces for living or sleeping, and serving spaces like corridors, wet cells (bathrooms or kitchen), and storage are required.
2. The building level inputs will influence mainly the material volumes of support (load-bearing) and infill (non-load-bearing) systems. Eight different structural systems with different material combinations are determined and embedded, and the users have a choice of the two types of infill system: partitions and frames that they would implement in the design with three options each.

There are several considerations when choosing the regression model for this particular use case:

1. The model should be able to act as a non-linear regressor - The interactions between inputs and outputs are non-linear. i.e. the number of rooms and the material volume needed for their enclosures is not simply a linear relationship.
2. The model should be able to gauge uncertainty of an output - For the model to be used as a decision-making tool, understanding the model uncertainty will add another dimension to obtain the optimal choice. Instead of a deterministic approach where model parameters have just a single value, models could take a stochastic approach and consider model parameters as having a probabilistic distribution with a range of values.
3. The model should be able to handle categorical and continuous variables - Certain inputs like construction and material systems are categorical (takes a fixed number of possible values) while building floor area is continuous (infinite number of values between any two values).

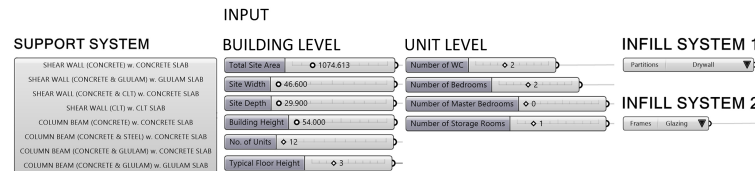


Figure 3. User input for both building and unit level in grasshopper.

One of the regression methods that fulfill those criteria is a Bayesian Neural Network (BNN), as proposed by (Gal et. al., 2016). By incorporating “stochastic dropout” within a neural network structure, a BNN is able to capture the uncertain spread of GWP resulting from generic design parameters across multiple building projects. Also, BNN regression approaches are able to scale with more input data due to its ability to detect all possible interactions between the input generic quantitative parameters and output GWP, as compared to more traditional regression methods like linear regression (Tu 1996). We assume that the building database available for modeling will grow in the future.

Data regarding generic quantitative parameters, eventual GWP were collated from 17 buildings in Section 3 (Data Collection), forming 17 data vectors. The first 14 data vectors were used for training the BNN, before the last three, which were extracted from a floor plan that is structurally verified, were used for validation. We utilized two BNNs for prediction, the building topology related information and construction system related information. The structure of BNN and their inputs and outputs are shown in figure 4. Categorical features in the data, like the ‘type of construction system’, ‘material for construction’, ‘sub-material for construction’ and ‘floor slab material’ exist as text. To allow the BNN to understand such forms of data, categorical data are converted to a mathematical vector, in a process known as “One-Hot Encoding.”

The building topological model statistically relates the basic building parameters to the functional spaces of a building, from which the ratio of circulation, served, and serving space of a building could be calculated. Similarly, the construction system model statistically relates the basic building dimensions, to the material volume required of the support and infill systems within a building, via an implicit understanding of the inner/outer enclosures of different areas within a building. The eventual GWP of the building is calculated from a multiplication of the material volumes with a material database.

Due to the presence of non-linear relationship between inputs and outputs, BNN should be more accurate than Linear Regression in modeling GWP. To compare the accuracy of the two types of regression approaches, the BNNs were benchmarked against the more commonly used Linear Regression method (Gadezi et al. 2016, Victoria et al. 2018), by calculating the Mean Absolute Percentage Error (MAPE) of the predictions made on the three validation data points, and evaluating the Cosine Similarity. In Figure 5, we show that for all the prediction outputs, the MAPE for linear regression is much higher than that of the BNNs.

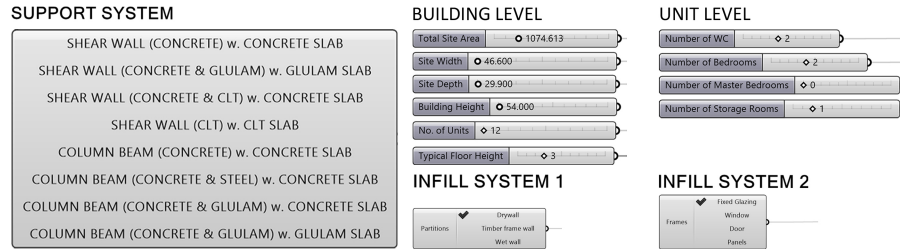


Figure 4. BNN Structure and Outputs .

For Cosine Similarity, the closer the values to 1, the better the performance. BNNs achieve an averaged Cosine Similarity score of 0.9893 across the three validation data points, compared to that of 0.9393 for the LR. As seen from the inhouse benchmarking test, BNN has better accuracy compared to linear regression and is able to provide an uncertainty indication when data is sparse. To provide a link to the visualization tool in Grasshopper, the trained BNN models are uploaded to the web and queried through a customized Python script in Grasshopper. Users without knowledge of Grasshopper will still be able to access the tool as an online resource.

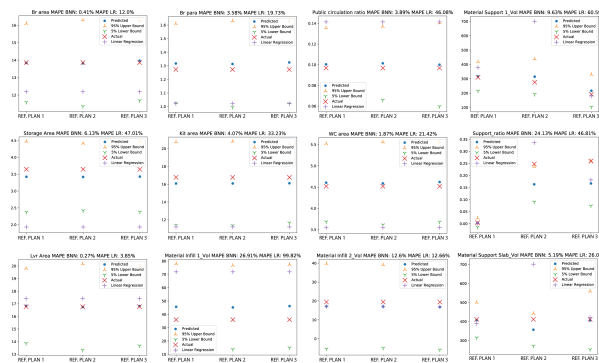


Figure 5. Error Comparison between BNN and Linear Regression for different projects .

5. Visualization in Tool

The tool provides a guided interface for users to input basic design parameters, as shown in figure 3, within Grasshopper or as an online resource. In return, the user will receive a ‘shoebox’ model generated from the input parameters, and a radar chart showing the different GWPs of different construction systems applied to the ‘shoebox’ model. The break-down of the different materials and their volume used in different construction systems are shown in a bar chart diagram in figure 6. In the background of the visualization tool, the basic design parameters will be passed to the BNNs and two types of outputs are produced: building topology related and construction system related output.

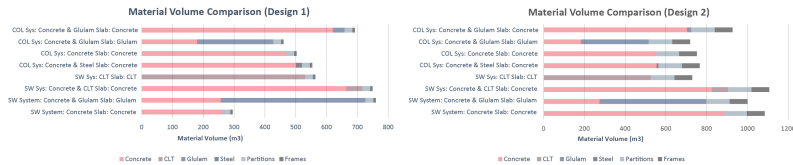


Figure 6. Bar chart shows a breakdown of the material combinations adopted in each identified construction system for both design variants. .

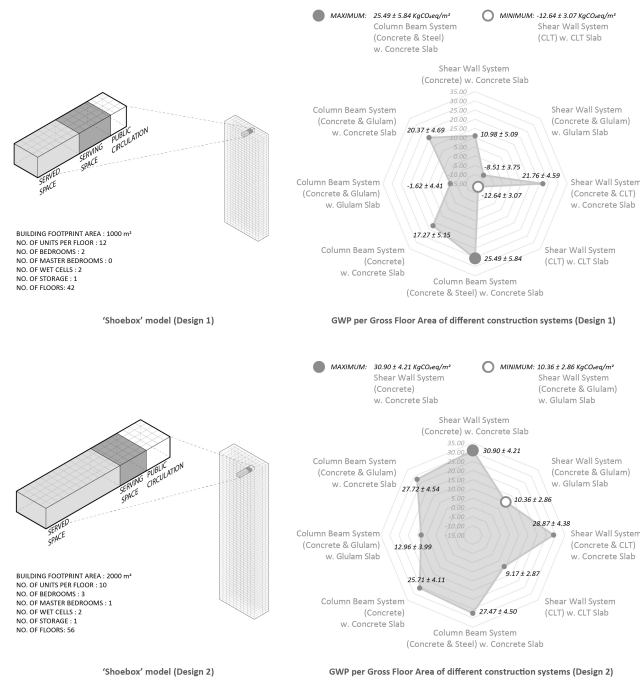


Figure 7. A 'shoebox' model is generated from two different input parameters, and a radar chart showing the different GWPs of different construction systems.

The 'shoebox' model, a three-dimensional representation of a building's primary spatial units, arises from the building topology related outputs of the BNN models. These outputs consist of the area of wet cells, bedrooms, master bedrooms, storage, living room, kitchen, and public circulation ratio. Multiplied by the number of each type of unit, these areas are tabulated into three essential spatial entities: served, and serving spaces within the unit, and unitized public circulation area, according to the definition by Louis Kahn (Sutanudjaja and Khan 2013). Wet cells, storage and kitchen are considered serving spaces while bedrooms, master bedrooms, and living rooms are considered served spaces. The 'shoebox' showcases the implicit relation of the three essential spaces in affecting

the material volume required for different construction systems, via the change in basic design parameters. Figure 7 shows the ‘shoebox’ visualization of two design variants with different inputs. Specifically for served space, an increase in the number of rooms in a unit would result in an increase in the number of enclosures, hence increasing in GWP value as illustrated in design 2.

The radar chart (figure 7) and the bar chart (figure 6) diagram are results of the construction system related output of the BNNs. The construction system consists of two types, shear wall or column beam system respectively, both with different material combinations, for a total of 8 different combinations. For the basic inputs given in figure 3, the eight different construction system combinations are tried to obtain eight different combinations of material volumes together with their uncertainty values. GWP values per gross floor area of the building are calculated according to the corresponding material volumes. Figure 7 shows two design variants with different building footprint and apartment units. The left image illustrates the three ratios of the spaces generated from the building topology related output. The right image demonstrates the different GWP values of different systems, denoting the maximum and minimum values.

6. Conclusion and Future Work

A predictive tool was developed to evaluate embodied carbon for early design iterations for residential buildings based on generic quantitative design parameters. Planners and designers can access the tool in Grasshopper or as an online resource, enter the relevant building information and obtain preliminary evaluations on different construction systems and material combinations in a simple visualization tool, represented in a ‘shoebox’. Since the predictive model is based on a Bayesian Neural Network, uncertainties could be measured, giving the users an extra dimension to ponder upon while making decisions with the tool. Also, more complex, non-linear interactions between input values and output values could be captured, as shown when the Bayesian Neural Network outperforms the Linear Regression method in our validation process.

A limitation of the current stage is that it is only drawing information from a small dataset, with 17 distinct residential building projects and limited material choices and construction systems. In order to further support the development of the tool, an open source database on building projects should be developed and training of the predictive model made into a web application. Users will be able to upload IFC files of recent building projects, that could be parsed and archived into the database. This will allow the tool to keep up-to-date with construction technology, and to scale beyond predicting GWP for residential building types with limited materials, to predicting additional LCA factors for a multitude of building types, with more detailed construction materials. Various stakeholders, from planners to experts to developers and policymakers could then select projects for the predictive module to reference, and download a customized predictive tool for their use case.

Nevertheless, this study underlines the potential role of data science in simplifying the LCA process.

Acknowledgements: The presented work is part of the research project ‘Next Residential High-rise’ funded by the SUTD-MIT International Design Centre (Project No. IDG21800103). Aspects relating to the component volume and GWP calculations were supported by Arup through a collaboration based on the Arup Global Research Challenge 2019. We would like to thank Zackary Xuereb Conti, Ray Cheng, Jun Wen Loo (all researchers at SUTD), Colin Yip, Xavier Zhou, Shiu Chern Ong (all Arup Singapore Pte Ltd) for their support.

References

- Basic, S., Hollberg, A., Galimshina, A. and Habert, G.: 2019, A design integrated parametric tool for real-time Life Cycle Assessment–Bombyx project, *IOP Conference Series: Earth and Environmental Science*.
- Dean, B., Dulac, J., Petrichenko, K. and Graham, P.: 2017, Towards zero-emission efficient and resilient buildings.: Global Status Report, *Global Alliance for Buildings and Construction (GABC)*, -, -.
- Ferlito, S., Atrigna, M., Graditi, G., De Vito, S., Salvato, M., Buonanno, A. and Di Francia, G.: 2015, Predictive models for building’s energy consumption: An Artificial Neural Network (ANN) approach, *2015 xviii aisem annual conference*, 1-4.
- Friedman, J., Hastie, T. and Tibshirani, R.: 2001, *The elements of statistical learning*, Springer series in statistics New York.
- Gadezi, S.S.S., Shafiq, N., Zawawi, N., Khamidi, M.F. and Farhan, S.A.: 2016, A multivariable regression tool for embodied carbon footprint prediction in housing habitat, *Habitat International*, **53**, 292-300.
- Gal, Y. and Ghahramani, Z.: 2016, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, *international conference on machine learning*, 1050-1059.
- Heckmann, O. and Schneider, F.: 2017, *Floor plan manual housing*, Birkhäuser.
- Kaufmann, H., Krötsch, S. and Winter, S.: 2018, *Manual of Multi-Storey Timber Construction*, Detail.
- Liu, H., Yang, Y.L., Alhalawani, S. and Mitra, N.J.: 2013, Constraint-aware interior layout exploration for pre-cast concrete-based buildings, *The Visual Computer*, **29**(6-8), 663-673.
- MacLeamy, P.: 2004, Collaboration, integrated information and the project lifecycle in building design, construction and operation, *The Construction Users Roundtable*, **12**, 12.
- Meex, E., Hollberg, A., Knapen, E., Hildebrand, L. and Verbeeck, G.: 2018, Requirements for applying LCA-based environmental impact assessment tools in the early stages of building design, *Building and Environment*, **133**, 228-236.
- Otovic, A., Jensen, L.B. and Negendahl, K.: 2016, Expansion in Number of Parameters-Simulation of Energy and Indoor Climate in Combination with LCA, *2016 ASHRAE Annual Conference*.
- Sinclair, D.: 2013, RIBA Plan of Work 2013 overview, *London: Royal Institute of British Architects*, -, -.
- Standardization, I.O.f.: 2006, *Environmental Management: Life Cycle Assessment; Principles and Framework*, ISO.
- Sutanudjaja, E.: 2013, Louis I Khan, Decorum In Architecture, The Idea of Modern Decorum and Human Agreement., -, -, -.
- Treloar, G., Love, P., Faniran, O. and Iyer-Raniga, U.: 2000, A hybrid life cycle assessment method for construction, *Construction Management & Economics*, **18**(1), 5-9.
- Tu, J.V.: 1996, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes., *Journal of clinical epidemiology*, **49**(11), 1225-1231..
- Victoria, M.F. and Perera, S.: 2018, Parametric embodied carbon prediction model for early stage estimating, *Energy and Buildings*, **168**, 106-119.