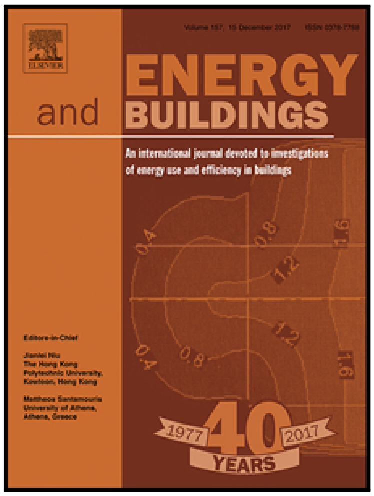
AUTHOR(S): VICTORIA, M.F. and PERERA, S.

TITLE: Parametric embodied carbon prediction model for early stage estimating.

YEAR: 2018

Accepted Manuscript



Parametric embodied carbon prediction model for early stage estimating

Michele Florencia Victoria , Srinath Perera

PII: S0378-7788(17)32181-3
DOI: 10.1016/j.enbuild.2018.02.044
Reference: ENB 8374

To appear in: *Energy & Buildings*

Received date: 29 June 2017
Revised date: 26 January 2018
Accepted date: 21 February 2018

Please cite this article as: Michele Florencia Victoria , Srinath Perera , Parametric embodied carbon prediction model for early stage estimating , *Energy & Buildings* (2018), doi: 10.1016/j.enbuild.2018.02.044

Parametric embodied carbon prediction model for early stage estimating

Dr. Michele Florencia Victoria (first author) The Scott Sutherland School of Architecture and Built Environment, Robert Gordon University, UK

Prof. Srinath Perera

School of Computing, Engineering and Mathematics, Western Sydney University, Australia

The focus of carbon management is shifting from operational carbon to embodied carbon as a result of the improved operational energy efficiency of buildings. Measuring and managing embodied carbon right from the early stages of projects will unlock a range of opportunities to achieve maximum reduction of emissions which could not be achieved otherwise during the latter stages. However, measuring embodied carbon during the early stages of design is challenging and highly uncertain due to the availability of limited design information. Therefore, the research presented in this paper addresses this problem in a structured and an objective way. A parametric embodied carbon prediction model was developed using regression analysis to estimate embodied carbon when only minimal design information is available and with less uncertainty. The model was developed by collecting historical data of office buildings in the UK from four different data sources and estimating embodied carbon by combining several estimating techniques. Wall to floor ratio and the number of basements were identified as the model predictors with a model fit of 48.1% (R^2). A five-fold cross-validation ensured that the model predicts within the acceptable accuracy range for new data. The developed model had and accuracy of ±89.35% which is within the acceptable for an early stage prediction model. In addition, the need for standardising embodied carbon measurements and to develop embodied carbon benchmarks to facilitate embodied carbon estimating throughout the project lifecycle was identified.

Keywords: Embodied Carbon, Finishes Index, Morphological parameters, Office Buildings, Regression Analysis, and Services Index.

Table 1: Summary of notations used

1. Introduction

Carbon management of buildings is imperative to achieve the emission reduction targets imposed on the built environment as the global construction industry is responsible for approximately 30% of the Greenhouse Gas (GHG) emissions (UK-GBC, 2014b, Olivier et al., 2016). Carbon management of buildings involves both operational and embodied carbon though embodied carbon is not regulated at present. However, the Green Construction Board (2013) of the UK suggests that 21% reduction of embodied carbon by 2025 and 39% reduction by 2050 have to be achieved for the UK to achieve its 50% and 80% of the overall reduction targets by 2025 and 2050 (The Green Construction Board, 2013). This echoes the need for regulating embodied carbon of buildings and calls for effective embodied carbon control mechanisms for the built environment.

Controlling embodied carbon requires carbon measurement in the first place. However, embodied carbon estimating is not a mature process as opposed to the operational carbon estimating practices. The work of Dixit et al. (2010, 2012) and De wolf et al. (2017) echoes the need for standardising embodied carbon measurement as there is a huge variation in the embodied carbon figures reported in the literature attributable to the variability of the assumptions made in the measurements. Embodied carbon can be calculated from raw material extraction (which is called the 'cradle') until the demolition of a building project (which is called the 'grave'). In some cases, end of life benefits resulting from reuse, recycle and recovery of building materials are accounted in the embodied carbon calculations (which is called as cradle-to-cradle). The scope of the embodied carbon calculation is called the 'system boundary'. Even though embodied carbon estimating practices are yet evolving, lessons can be learned from the well-developed cost estimating practices (Ashworth and Perera, 2015, Perera and Victoria, 2017), as both cost and carbon can be estimated concurrently due to the same determinants (material, labour (only for cost) and plant). Accordingly, it is proven in the cost studies that the highest reduction potential can be achieved during the early stages of design (Asiedu and Gu, 1998) and RICS (2014) suggests that the same is true in the case of embodied carbon.

There are a range of case studies reported in the literature on embodied carbon of buildings and the reduction potential of embodied carbon have been explored through alternative design solutions. Figure 1 summarises the embodied carbon values of different types of buildings obtained from various studies. It should be noted that the values reported only include the embodied carbon of the building structure. The values of semi-detached houses were obtained from Hacker et al. (2008) and Monahan and Powell (2011). A two storeyed semi-detached house was studied in both cases and alternative structural options were simulated to analyse the impact of design decisions on the embodied carbon of the building. Studies demonstrate that the EC of the structure of the case study building ranges from 355 $kgCO_2/m^2$ to 569 $kgCO_2/m^2$ and concluded that the embodied carbon can be reduced by 51% from the structure of the building alone. The embodied carbon values of other types of buildings were obtained from the study conducted by Sansom and Pope (2012). Single case studies were employed for each type of building and the impact of alternative structural forms on the embodied carbon of each building was studied. Further, Sansom and Pope (2012) adopted a cradle-to-grave system boundary which includes the emissions associated with the raw material extraction up to the demolition of the building (however, the study excluded recurring embodied carbon which covers repair, maintenance and replacement during the use phase of the building). Estimating embodied carbon using a life cycle model is a more holistic approach and desirable as it helps to see the macro picture of the emission and cost savings achievable during the life cycle of the building. For instance, Kneifel (2010) showed that energy efficient technologies can reduce the energy use in commercial buildings of up to 40% at a negative life cycle cost and suggest that initial investments on energy efficient technologies pays back several folds in the long run. However, life cycle assessments are challenging and it is hugely influenced by project specific factors.

Figure 1: Embodied carbon values of different types of buildings from the literature

Embodied carbon analyses of office buildings are presented separately in Figure 2 due to its popularity in the academic literature. Findings of four studies are mapped onto a spider web diagram to demonstrate the variation in the embodied carbon values of office buildings. Clark (2013) reported embodied carbon analyses of office buildings ranging from low to high rise buildings, structure only analyses to whole building analyses and cradle-to-gate analyses to cradle-to-grave analyses. Hence, the reported embodied carbon values ranges from 300 $kgCO_2/m^2$ to 1,650 $kgCO_2/m^2$. On the other hand, Sansom and Pope (2012) adopted a cradle-to-grave system boundary excluding recurring embodied emissions similar to other building types presented in Figure1. The change in the embodied carbon influenced by the change in the structural form of the selected office building was investigated by Sansom and Pope (2012). Hence, the variation is small and it was shown that 11% reduction in the embodied carbon is achievable (structure only) in that particular building. Victoria et al. (2015b) reported cradle-to-gate embodied carbon analyses of seven office buildings which range from 271 $kgCO_2/m^2$ to 706 $kgCO_2/m^2$. However, these embodied carbon analyses exclude some of the major building services, hence not holistic. Halcrow Yolles (2010b) studied three low-rise office buildings within a cradle-to-gate system boundary. The embodied carbon of the three office buildings ranges from 538 $kgCO_2/m^2$ to 924 $kgCO_2/m^2$ (excluding major building services). Further, Halcrow Yolles (2010b) found that improvement to the operational energy can escalate the embodied carbon up to 25% (Halcrow Yolles, 2010b).

Figure 2: Embodied carbon studies on office buildings

Even though the reported values in the literature are non-comparable due to the difference in the scope of studies, EC estimating practices enable design and construction professionals to make informed decisions. Therefore, it is beneficial to review past studies to capture existing embodied carbon estimating practices at different stages of projects.

Table 2 provides an overview of embodied carbon estimating practices and data sources employed at various stages of construction projects. Accordingly, embodied carbon estimating practices are prevalent at detailed stages of design (3-Developed Design and 4-Technical Design) compared to early stages (3-Concept Design). However, EC estimating practices should be harnessed during the early stages of design to exploit the maximum emission reduction ability of buildings (RICS, 2014, Victoria et al., 2015a). Even though 'Construction Carbon Calculator' can be used during early stages of design, it lacks transparency of the underlying methodology. This questions the scientific validity of the tool. This was identified as a gap and a method is proposed in this paper to develop an embodied carbon prediction model to facilitate EC estimating, particularly at the concept design stage, which was inspired by the design economics literature (Seeley, 1996, Ashworth and Perera, 2015, Dell'Isola and Kirk, 1981, Collier, 1984, Robinson and Symonds, 2015). The proposed method is verified by developing a model by collecting data and testing the model for its accuracy in prediction which proves the scientific validity of the findings. In addition, the model can also be developed into a scalable decision support system by integrating other dimensions of construction projects such as embodied energy, waste, time and cost to paint a holistic picture as conceptualised by Abanda et al. (2013) given that the identified limitations are addressed reasonably.

Table 2: A review of embodied carbon estimating practices adopted in past studies

2. Methodology

2.1. Research scope

Non-domestic buildings are responsible for higher embodied carbon emissions compared to domestic buildings and infrastructure (The Green Construction Board, 2013). It is also predicted that non-domestic floor area is expected to increase by 35% in the UK by 2050 (UK-GBC, 2014a). Hence, the focus of the study was confined to non-domestic buildings. In particular, office buildings are expected to grow at a rate of 2.7% which is higher than the other types of non-domestic buildings (The Green Construction Board, 2013). Further, The Green Construction Board (2013) states that the commercial office buildings are superior to other types of building in terms of the clarity of the definition and availability of data which eliminates the risk of uncertainty in modelling. In addition to that office buildings are the key focus of many scholars and an extensive amount of work has been undertaken to improve the energy efficiency of office buildings (Halcrow Yolles, 2010a, Yohanis and Norton, 2002, Halcrow Yolles, 2010b, Cole and Kernan, 1996, Wu et al., 2012). Due to these reasons, office buildings were selected as the scope of the study.

Further, the system boundary of the embodied carbon analysis was limited to 'Cradle-to-Gate' due to the embodied carbon inventory used and the unavailability of project specific data. Process based estimating method was used to estimate the embodied carbon of buildings using the quantity of materials or items and the carbon emission factors of materials or items. The equation used to calculate the embodied carbon is presented in Equation 1.

Equation 1: Formula to estimate embodied carbon of materials or items of buildings

$$EC_{m/i} = Q_{m/i} . ECF_{m/i}$$

Where, $EC_{m/i}$ refers to the total embodied carbon of a particular material or an item in a building, $Q_{m/i}$ is the total quantity of the respective material or item and $ECF_{m/i}$ is the embodied carbon factor of the respective material or item.

In addition, the term early stage refers to the first three stages of the Royal Institute of British Architects (RIBA) plan of work 2013 namely, strategic brief, preparation and brief and concept design (RIBA, 2013). In particular, the developed model cater the estimating need of the 2-Concept Design stage.

2.2. Overview of the Method

The research method involved both quantitative and qualitative data collection and analysis techniques and consists of several steps of data processing which is illustrated in Figure 3. All these steps are further explained in the subsequent sections of the methodology. Step one involved the preparation of EC estimates of Dataset 1 using the EC factors from supporting data. Step 2 was the process of deriving the elemental EC rates which will be referred to as 'Embodied Carbon Element Unit Rates' (EC-EURs) in the paper. EC-EUR of an element is the carbon embodied in one unit of the element considered and can be denoted as follows:

Equation 2: Formula to calculate EC-EUR

$$EC - EUR_i = \frac{EC_i}{EUQ_i}$$

Where, $EC - EUR_i$ is the Embodied Carbon Element Unit Rate of element '*i*', EC_i is the embodied carbon of element '*i*' and EUQ_i is the Element Unit Quantity of element '*i*'.
EUQ of a building element is calculated in accordance with the guidance provided in the New Rules of Measurement (NRM) documents (RICS, 2012). For instance, EC-EUR of the Substructure can be obtained by dividing the Substructure EC by the Element Unit Quantity (EUQ) of the Substructure, where, Substructure EUQ according to the NRM is the building footprint area.

Figure 3: Overview of the methodology

These developed EC-EURs along with supporting data were used to prepare EC estimates of buildings in Dataset 4 as it contained only EUQs and brief element specifications which made bottom-up approach to estimating non-applicable for certain elements. Such use of alternative techniques to bottom-up approach to estimating EC were evidenced in past studies (Cole and Kernan, 1996, Monahan and Powell, 2011. However, as the next step (Step 4), an independent EC dataset 'Dataset 3' was used to validate Dataset 4 to vindicate the method adopted.

Meanwhile, Finishes Quality Indices and Services Quality Indices were developed using qualitative data collection and analysis techniques as finishes quality and services quality were identified as two key variables affecting cost, hence, most likely to influence EC of buildings too. This created the need to develop quantitative indices for finishes and services quality levels of office buildings to include as ordinal variables in the regression analysis. Consequently, finishes index was developed from an expert forum and the services index was developed from a document review which is denoted by Step 5. This lead to the tagging of the finishes and services quality levels of buildings in Dataset 4 in a uniform way.

Step 6 involved the multiple regression analysis of Dataset 4 with the selected design variables and the regression assumptions were tested for its validity (Step 7). In Step 8, a five-fold cross validation was employed to measure the prediction performance of the model with internal and external data to assess the generalisability of the model. The final model was validated in Step 9 and its prediction performance was analysed.

2.3. Data collection and processing

Historical project data were collected from four different sources due to the unavailability of a standalone embodied carbon database. Figure 4 presents three main categories of data obtained including primary, secondary and supporting data. Bill of Quantities (BOQs) or detailed cost plans and layout drawings were obtained from thirteen (13) Quantity Surveying (QS) practices which forms the primary data of the study. Secondary data were obtained from public online databases and a special QS database which constitute of embodied carbon and cost analyses of buildings. Supporting data are the published cost and carbon data books/inventories which are used in conjunction with the collected primary and secondary data to build up the cost and carbon estimates. These include: Inventory of Energy and Carbon (ICE) (Hammond and Jones, 2011), the UK building Blackbook (Franklin & Andrews, 2011) and manufacturer specific data. Each dataset obtained was mapped against the data requirement of the research which is shown in Table 3. Accordingly, none of the dataset met the complete data requirement of the research. Hence, a statistically significant sample for the study was developed by obtaining cost analyses from Building Cost Information Services (BCIS), an online cost database (RICS, 2016). All the available data that fulfilled the data requirement of the study was obtained from BCIS which resulted in 41 buildings with specification, element quantity and cost information. Embodied carbon estimates were produced for the building data obtained from BCIS using the embodied carbon data collected from other datasets using a range of estimating techniques such as bottom-up approach, statistical averages and extrapolation. This method of filling gaps in the data was inspired from past studies which used embodied energy/carbon values and averages from published studies to complete the embodied energy/carbon analysis (see, Cole and Kernan, 1996, Monahan and Powell, 2011).

Figure 4: Types and sources of data obtained (numbers within the brackets denote the sample size)

Table 3: Mapping each dataset against the data requirement

Figure 5 illustrates the process involved in the development and validation of Dataset 4 which is the final study sample consisting of 41 office buildings. The developed sample (Dataset 4) was complemented by Dataset 1 and Dataset 2 and for that reason it was validated by an independent dataset (Dataset 3). Data inadequacies noted in Dataset 4 disqualifies the applicability of bottom-up approach to estimating to all building elements. Especially, elements such as Substructure, Frame, Upper Floors, Roof, Fittings, Furnishings and Equipment and Services are measured in m^2 (while its components measured differently such as m^3, m, tonnes, numbers etc.) and lack detailed specification of its sub-elements. Therefore, Embodied Carbon Element Unit Rates (EC-EURs) were developed from Dataset 1 and Dataset 2 to assist in the EC estimating of Dataset 4.

Figure 5: Development and validation of Dataset 4

Accordingly, EC-EURs of different types of Substructure, Frame, Upper Floors, and Roof were obtained from Dataset 1 and extrapolation method was used to derive the embodied carbon rates of these elements of Dataset 4 using the cost of the elements presented in the BCIS. Similarly, EC-EURs of Fittings, Furnishings and Equipment and Services were obtained from Dataset 2 and the EC of the rest of the elements were estimated from a bottom-up approach using the UK Building Blackbook, ICE and manufactures' embodied carbon data. The developed Dataset 4 was validated using Dataset 3 (WRAP and UK-GBC, 2014) to ensure the reliability of Dataset 4 as it was developed from various sources. In doing so, embodied carbon data of Dataset 4 had to be grouped into six categories to be comparable with the EC data of Dataset 3. These groups include: Substructure, Superstructure Structural, Superstructure Non-Structural, Envelope, Internal Finishes and External Works. However, only four categories were able to be verified, because: (1) Envelope embodied carbon was not available for Dataset 3 and (2) External works were excluded from embodied carbon estimates due to its project and client specific nature and it depends on the topography and shape of the site. Table 4 presents the sub-elements of the validated four element categories as prescribed in Dataset 3.

Table 4: Element groups as prescribed in WRAP dataset (Dataset 3)

Table 5 presents the descriptive statistics of the two independent datasets. Kolmogorov-Simirnov test was used to ascertain the normality of the datasets within the selected element groups to select the appropriate test to compare the embodied carbon data in the two groups. According to the Kolmogorov-Simirnov test statistics, only the embodied carbon data of Superstructure-Structural group conformed to a normal distribution. Hence, log transformation was applied to the remaining groups in an attempt to achieve normality and the Kolmogorov-Simirnov test was repeated. Log transformations resulted in both Substructure and Superstructure Non-Structural data groups conform to normal distributions. Hence, a two sample independent t-Test was conducted within these groups while Mann-Whitney U test, which is a non-parametric equivalent of the independent sample t-test was, conducted for Internal Finishes group as one of the datasets within the group did not comply with the normality assumption.

Descriptive statistic presented in Table 5 suggests that the means of all elemental groups are almost similar except for Superstructure-Structural. This is again confirmed by t-Test statistics which suggests with 95% confidence that there is no sufficient evidence to say that the means of the (log of) Substructure and (log of) Superstructure Non-Structural of the two samples are significantly different while there is sufficient evidence (sig. < 0.05) to conclude that there is a significant difference between the mean embodied carbon values of Superstructure Structural of the two samples (see,

Table 6 for t-Test statistics). Possible reason for this difference could be attributable to Roof embodied carbon as it could involve a range of alternative and complex specifications which is unknown for Dataset 3. Similarly, embodied carbon of Upper Floors could have also influenced the identified difference if the buildings in Dataset 3 have predominantly timber floors and pre-cast floors.

Table 5: Group statistics for individual element categories

Table 6: t-Test statistics of the two samples – Dataset 3 and Dataset 4

On the other hand, Mann-Whitney U test confirmed that the means of the two datasets within the Internal Finishes group are not significantly different (Sig.=0.449). This interpretation is true only when the two groups follow the same distribution which is also verified in this case (Sig.=0.627). Hence, it was concluded that the embodied carbon estimates of Dataset 4 are reliable though there is ambiguity concerning the estimate of Superstructure Structural. This could not be further investigated due to the lack of specification information of buildings in Dataset 3. This is identified as a limitation of the study.

2.4. Data analysis techniques

2.4.1. Finishes index development

An objective index for the finishes quality of office buildings was developed to use finishes quality as a predictor variable in the developed model. The process followed in the development of the finishes quality index is presented in Figure 6. Initially, a conceptual finishes quality index was developed by surveying common types of wall, floor and ceiling finishes in office buildings and classifying them into three quality categories namely Basic, Moderate and Luxury. The conceptual finishes index was then verified through a Delphi based expert forum consisted of five experts. Clayton (1997) state that Delphi technique is appropriate when seeking the consensus of experts on content validity and it allows rigorous and systematic data collection and dissemination without the need of the experts to travel and meet as a group at a particular time and a place. Construction professionals with more than ten (10) years of industry experience and with a Chartered membership were chosen purposively to be the experts of the panel as RICS (2009) stipulates these two criteria for a person to be considered as an expert. Further, a panel size of 5 to 10 is suggested for a heterogeneous population (Clayton, 1997). Accordingly, four (4) QSs and an Architect was selected as the experts of the panel. The experts were given the opportunity to re-evaluate their responses in the second round and the final finishes index was derived as the consensus was reached in the second round. The proposed finished quality index is presented in Table 7.

The quality level of the wall, floor and ceiling finishes were ascertained using a weighted average method and the overall finishes quality index of the building was calculated using the formula presented in Equation 3.

Equation 3: Formula to calculate the overall finishes quality index of the building

$$FI_{Building} = \sum_{Wall}^{Ceiling} A_i . FI_{W,F,C}$$

Where, $FI_{Building}$ denotes to the overall finishes quality of the building, A_i is the area of wall/floor/ceiling finishes as a percentage of the total finished area, $FI_{W,F,C}$ is the overall wall/floor/ceiling finishes index. Finishes index for wall/floor/ceiling finishes are calculated as follows:

$$FI_{W,F,C} = \sum_{Basic}^{Luxury} a_i . I_i$$

Where, a_i is the area of basic/moderate/luxury finishes as a percentage of the total wall/floor/ceiling finishes area, I_i is the respective index assigned for basic/moderate/luxury quality of finishes.

Figure 6: The development process of the finishes quality index

Table 7: The proposed finishes quality index for the study

2.4.2. Services index development

A range of price books were reviewed to adapt an objective service quality level for the study and the developed service quality index was adapted from Spon's Mechanical and electrical Services Price Book (Davis Langdon Consultancy, 2014) which is presented in Table 8. Essential building services include sanitary appliance, water installations, disposal installations, space heating systems, ventilation systems, electrical installations, fire and lighting protection, communication and security installations and automated buildings refer to the buildings equipped with a Building Management System (BMS).

Table 8: The proposed service quality index

2.4.3. Model development

The process of formulating the embodied carbon prediction model follows the basic structure of cost modelling research which involves three main stages including conceptualization of the model, model formulation (by collecting data) and validation of the model (Ashworth and Perera, 2015). Multiple regression analysis was selected over other modelling techniques due to its well defined mathematical basis, transparency and its popularity within the construction management discipline (Karshenas, 1984, Kouskoulas and Koehn, 2005, Karanci, 2010, Kim et al., 2004, Alshamrani, 2016). The conceptual model is presented in Equation 4.

Equation 4: The conceptual embodied carbon prediction model

$$\hat{y} = a_0 + a_1 x_{W:F} + a_2 x_{ASH} + a_3 x_{BH} + \cdots + a_{FI} + a_{SI}$$

Where,

$\hat{y}$ $- Estimated\ Embodied\ Carbon\ per\ GIFA\ of\ the\ building$

a_0 $- Regression\ constant$

a_1 $- Regression\ coefficient\ of\ x_{W:F}$

$x_{W:F}$ $- Wall\ to\ Floor\ ratio\ of\ the\ building$

a_2 $- Regression\ coefficient\ of\ x_{ASH}$

x_{ASH} $- Average\ Storey\ Height\ of\ the\ building$

a_3 $- Regression\ coefficient\ of\ x_{BH}$

x_{BH} $- Building\ Height$

a_{FI} $- Finishes\ Index\ of\ the\ building$

a_{SI} $- Services\ Index\ of\ the\ building$

Eventually, the model parameters were estimated through the regression analysis using a backward method. This method accommodates all input variables in the first run and eventually removes one variable at a time which is the least significant in the model and Field (2013) suggests that this is more accurate than forward and stepwise regression methods.. Afterwards, regression assumptions were tested as the model cannot be considered valid if key regression assumptions are violated. The regression assumptions and the mechanisms used to test these assumptions are listed below:

1. Normality of data of the dependent variable – descriptive statistics (skewness)

2. Linearity between dependent and independent variables - residual plot (residuals in the standardised residual plot should be randomly distributed)
3. No mulicollinearlity between independent variables –Variance Inflation Factor (VIF) of the model (VIF between 5 and 10 indicates high correlation and VIF beyond 10 reveals that the regression correlations are poorly estimated)
4. Residuals are homoscedastic - scatterplots of residuals against predicted values (residuals are expected to be randomly distributed and not demonstrate any patterns)
5. Residuals are not autocorrelated - the Durbin-Watson test statistics (d). See, Equation 5.

Equation 5: Durbin-Watson test statistics

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

Where, n is the number of observations and e_i is the residual of the i^{th} observation. d = 2 indicates no autocorrelation as the value of d will always lies between 0 and 4. d is compared to the lower and the upper critical values ($d_{L,\alpha}$ and $d_{U,\alpha}$) at significance α (See Table in Appendix 3 for critical values of the Durbin Watson Score).

Finally, the developed model was tested for its accuracy of predictions. The coefficient of Variation (CV) is the metric used to check the accuracy of predictions of the model. CV is calculated as a percentage of the standard deviation of the residuals divided by the mean of the observed values of the dependent variable, which is presented in Equation 6.

Equation 6: Formula to calculate coefficient of variation

$$CV = \frac{\sqrt{\frac{\sum(e_{i-}\,\overline{e})^2}{n-1}}}{\overline{y}} X\ 100$$

Where, e_i is the residual, $\overline{e}$ is the mean of the residuals, n is the number of observations, $\overline{y}$ is the mean of the observed values (actual embodied carbon values).

For instance, a CV of 15% implies that the accuracy of the prediction of most of the cases (68%) would fall between ±85%. Hence, a smaller CV is desirable. However, Ashworth and Skitmore (1999) from a thorough analysis of the past studies suggested that CV of 15% to 20% of prediction accuracy is acceptable for early design stage cost estimates while Peurifoy and Oberlender (2002) proposed that an accuracy between +25% to -5% is acceptable for a conceptual estimate. However, a lower CV implies a better model prediction. Therefore, a prediction accuracy of CV ±20% is considered sufficient to validate the models. Nevertheless, it is said and proved that the CV of a model will deteriorate when the model tends to predict cases outside its database (McCaffer, 1999). Hence, the prediction accuracy of the model has to be assessed using internal (data forming the model) and external data (data outside the model) in order to assess the generalisability of the model. However, it is not sensible to leave some data out from the sample to test the model as the sample size of the study was small. Cross-validation is a widely accepted method to test the prediction performance of a model when the sample size is small. Kohavi (1995) noted that the variance is reduced in a k-fold cross validation with moderate k values (10-20) while variance is increased as k decreases (2-5) and the sample size gets smaller due to the instability of models. Hence, many scholars accept a ten-fold cross-validation as valid. However, the variance that was noted in the five-fold cross validation of the study models was very low, suggesting the formulated model is stable.

3. Results and discussion

3.1. Analysis of variables

Descriptive statistics and the correlation matrix of the variables were produced to investigate the normality aspect and multicollinearlity. Regression assumes that the dependent variable is normally distributed. In this case the dependent variable is the EC per GIFA. The statistics for skewness and kurtosis give an indication of the normality of the data distribution of the variables. Skewness of a normally distributed variable will have a value of 0. Miles and Shevlin (2001) suggest that there is little problem if the skewness statistics is less than 1.0 and skewness statistics between 1.0 and 2.0 is also cautiously acceptable attributing to the fact that it might have an impact on the estimates. However, skewness statistics above 2.0 indicates a serious problem with normality. According to the skewness statistics presented in Table 8, average height, wall to floor ratio, circulation ratio and embodied carbon per GIFA are less than 1.0 while GIFA and building height lies between 1.0 and 2.0, which ensures no major violation of the assumption of normality had occurred with the selected sample.

Table 9: Descriptive statistics of variables

On the other hand, correlation matrix was produced to detect multicollinearity between independent variables (see, Table 10). Regression works based on the assumption of correlation between dependent and independent variables only, hence, a strong correlation between independent variables will affect the validity of the model. Accordingly, a correlation coefficient of more than 0.7 between two independent variables signposts the presence of multicollinearity (Miles and Shevlin, 2001). However, the correlation matrix confirms that there is no multicollinealriy between any independent variables.

Table 10: Correlation matrix

3.2. Regression analysis

As discussed in the methodology section, backward method regression analysis was performed with the identified variables to derive the best predictive regression model. The summary of the model is presented in Table 11. The best predictive model was derived in the fifth step with wall to floor ratio and the number of basements being identified as the most significant predictor variables of EC per GIFA of office buildings. R^2 indicates the percentage change in the dependent variable explained by the independent variables in the model. Model summary displays that no much improvement is achieved in adjusted R^2 when progressing from one step to the other and the standard error of estimate also shows little improvement. However, a drastic drop from R^2 to adjusted R^2 is clearly notable in the first four steps while the drop in less in the fifth model. Further, 48.1% of the change in the dependent variable is explained by wall to floor ratio and number of basements in Model 5 while 48.8% and 49.5% of change is explained by services index and finishes index in Model 3 and Model 4, which is better than Model 5. However, finishes and services indices were found to be insignificant in the models (Sig. < 0.05). Therefore, Model 5 is selected as the best predictive EC per GIFA model.

Table 11: Summary of the models produced using the backward regression analysis

The EC per GIFA model is presented in Equation 7. Where, $\hat{y}$ is the estimated EC per GIFA of the building, $x_{W:F}$ is the wall to floor ratio and x_B is the number of basements. The model indicates that an increase in one unit of wall to floor ratio (say, 0.3 to 1.3) while holding the number of basements as constant will increase EC per GIFA by 164.08 $kgCO_2/m^2$ and adding a basement will increase EC per GIFA by 68.15 $kgCO_2/m^2$ for a given wall to floor ratio. Both coefficients are reasonable as a higher wall to floor ratio implies higher façade area and more basements implies more material and plant inputs increasing the EC per GIFA of a building. Further, it can be noticed that the constant is high compared to other coefficients. This can be explained by the descriptive statistics of the sample data presented in Table 8, as the EC per GIFA ranges from 551 $kgCO_2/m^2$ to 916 $kgCO_2/m^2$. Even the smallest building has an EC per GIFA value of 834 $kgCO_2/m^2$ GIFA. Therefore, it is clear from the coefficient that the minimum EC per GIFA of a building will be more than 530.62 $kgCO_2/m^2$ as per the findings.

Equation 7: EC per GIFA model

$$\hat{y} = 530.62 + 164.08x_{W:F} + 68.15x_B$$

However, it was surprising that the building height was identified as a significant predictor as the literature (Luo et al., 2015) suggest that building height (no. of storeys) and EC per GIFA has a strong positive correlation while the relationship found in the study was moderate (0.392 at the 0.05 level). Hence, it can be articulated that when fitting the regression model other variables (wall to floor ratio and basements) have overridden the building height. This may be due to the selected sample and with a different sample different result can be expected.

3.3. Regression assumptions

Normality assumption was tested by performing Kolmogorov Smirnov test on the regression residuals and the test results confirmed that the residuals are normally distributed (Kolmogorov Smirnov Test Statistic 0.077, Sig. 0.200). Linearity assumption was ratified by the residual plot which showcases a random spread of residuals (see, Figure 7). VIF of the variables in Model 5 was close to 1, which proves of no multicollinearity in the model. Scatterplots for standardised residuals of the regression presented in Figure 7 confirms that the residuals are homoscedastic (randomly distributed) and do not demonstrate any significant patterns. The Durbin Watson score of the model was 1.879 which is greater than $d_{U,\alpha}$ ($d_{U,\alpha,}$ =1.60) indicating no positive autocorrelation among the residuals. Similarly, 4-d (4 – 1.879 = 2.121) is also greater than $d_{U,\alpha}$ confirms no negative autocorrelation. Therefore, the model satisfies all necessary regression assumptions.

Figure 7: Scatterplot of standardised predicted value vs. standardised residuals of regression

3.4. Five-fold Cross-validation

A five-fold cross validation was performed to test the prediction performance of the model with different datasets, hence, to provide an assurance of the generalisability of the model. Data was split into five sets as opposed to ten sets (which is considered a sufficient number) because of the smaller sample size of the dataset. Accordingly the data was partitioned into five (5) sets, each set containing eight (8) buildings except for one set which contained nine (9) buildings (see, Figure 8). Regression analysis was performed with four sets leaving one set out at a time for testing the model. The process was iterated five time until each set has been left out for testing the model. CV was calculated for both training set and the test set in each fold. The results obtained at each fold are presented in Table 12. According to the cross validation results, all models performed alike except for Model 1 which had an additional predictor, Finishes Index. However, R^2 of Model 1 was the lowest of all which indicates overfitting of the model. In terms of the CV, all models performed well with internal (training set) and external (test set) data which is well within the acceptable CV range for an early stage prediction model. On average, the CV of the models for internal data and external data were calculated to be 10.30% and 11.91% respectively. Further, the prediction performance has not deteriorated drastically when predicting for new data which assures the generalisability of the model. Therefore, the derived model can be approved of being capable of predicting EC per GIFA of new data with an acceptable level of accuracy.

Figure 8: Five-fold cross validation data splitting of Dataset 4

Table 12: Summary of five-fold cross validation outcomes

3.5. Analysis of the Prediction Performance of the Final Model

The CV of the final model (Equation 7) was 10.65% and the model performance was analysed in terms of different storey cluster such as 1-2 storey, 3-5 storey and 6+ storeys as presented in Figure 9. The model prediction lies within the 20% margin for all buildings except for one building in the 3-5 storey cluster. However, the prediction performance for 6+ storeys cluster cannot be certainly ascertained as there is only one building in the sample which has 6 storeys. Further, the model seems to predict closer to the observed values in the 1-2 storey cluster in comparison to 3-5 storey clusters. In addition, the accuracy ranges from -19% to 20% with a CV of 10.4% in 1-2 storey cluster and -25% to 17% with a CV of 11.2% in 3-5 storey cluster. This implies that the model predicts at its best in the 1-2 storey cluster.

Figure 9: The model predictions at different storey clusters

4. Conclusions

Increasing significance of embodied carbon in buildings and challenges in estimating embodied carbon during early stages of design due to limited design information became the driver of the study. The lessons learned from cost modelling literature enabled to develop the idea of a parametric embodied carbon model to estimate embodied carbon of conceptual building designs using quantitative and qualitative design variables as predictors. Subsequently, historical project data of office buildings were collected from different sources and the embodied carbon was estimated using different estimating techniques such as bottom-up approach, statistical averages and extrapolation. Regression analysis was used to develop the embodied carbon that predicts the embodied carbon per GIFA using the basic design variables. Qualitative variables such as finishes quality and services quality of building were objectivised by developing separate quality indices. However, only two variables were identified as statistically significant predictor variables in the model which are wall to floor ratio and the number of basements. The outcome was surprising as it was anticipated that building height and quality were also expected to be key determinants of embodied carbon of buildings as in the case of cost. This may be due to the selected sample which consist of low to medium rise buildings up to 6 stories. The embodied carbon model explains 48.1% of the variation in EC per GIFA by wall to floor ratio and the number of basements. Further, the model has an average CV of 10.30% for internal data and 11.91% for external data which is an acceptable accuracy range for an early stage prediction model. However, the final model predicts at its best at the 1-2 storey cluster.

The model facilitates easier and faster prediction of embodied carbon during the early stages of design and allows comparisons between alternative design solutions. In addition, this model will encourage designers to rethink their designs right from the early stages of the project. However, it should be noted that the model is limited in its scope as it is developed solely for office building of low to medium rise in the UK. Hence, models for different types of buildings with different height categories in different parts of the world should be formulated by collecting location specific data. In addition, changes in the method of manufacturing of building materials have to be factored in embodied carbon estimates as the embodied carbon of materials is affected by processes. For instance, embodied carbon of materials are deemed lower if fossil fuels are substituted by renewable energy sources. Hence, such global variables need to be considered in future embodied carbon studies even though it is not given due consideration in current research which relies greatly on existing embodied carbon databases. In contrary, cost estimates are adjusted for time and location related variations using time and location indices.

Predicted embodied carbon covers a cradle-to-gate boundary, which implies that transport is excluded (other than raw material transport to factory gate). However, transport could be a significant component of the total embodied carbon of projects which use an extensive amount of imported materials. Therefore, the users of the model should be mindful of such anomalous circumstance and make necessary allowances in the estimate. Further, the proposed methodology can be adopted to formulate similar models for different types of buildings at different locations. Even though the model presented in this paper appears to be a manual model, it can be developed into a scalable decision support system with the use of a spreadsheet or advanced programming languages to make it more user friendly. Such a model can be elevated by integrating other dimensions of construction projects such as embodied energy, waste, time and cost to make more rational decisions underpinned by sustainable practices in the built environment.

References

ABANDA, F.H., TAH, J.H.M. & CHEUNG, F.K.T. 2013. Mathematical modelling of embodied energy, greenhouse gases, waste, timeecost parameters of building projects: A review. *Buildings and Environment*. 59, 23-27.

ALSHAMRANI, O. S. 2016. Construction cost prediction model for conventional and sustainable college buildings in North America. *Journal of Taibah University for Science*.

ASHWORTH, A. & PERERA, S. 2015. *Cost studies of buildings*, Oxon, Routledge.

ASIEDU, Y. & GU, P. 1998. Product life cycle cost analysis: State of the art review. *International Journal of Production Research*, 36(4)**,** 883-908.

CLARK, D. H. 2013. *What Colour is your Building?: Measuring and reducing the energy and carbon footprint of buildings*, UK, RIBA Enterprises Limited.

CLAYTON, M. J. 1997. Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, 17(4)**,** 373-386.

COLE, R. J. & KERNAN, P. C. 1996. Life-cycle energy use in office buildings. *Building and Enrironment*, 31(4)**,** 307-317.

DAVIS LANGDON CONSULTANCY 2014. *Spon's mechanical and electrical services price book 2014*, London, Taylor & Francis.

DE WOLF, C., POMPONI, F. AND MONCASTER, A. 2017. Measuring embodied carbon dioxide equivalent of buildings: A review and critique of current industry practice, *Energy and Buildings*, 140, 68-80.

DIXIT, M. K., FERNÁNDEZ-SOLÍS, J. L., LAVY, S. & CULP, C. H. 2010. Identification of parameters for EE measurement: A literature review. *Energy and Buildings*, 42(8), 1238-1247.

DIXIT, M. K., FERNÁNDEZ-SOLÍS, J. L., LAVY, S. & CULP, C. H. 2012. Need for an EE measurement protocol for buildings: A review paper. *Renewable and Sustainable Energy Reviews*, 16(6)**,** 3730-3743.

FIELD, A. P. 2013. *Discovering statistics using IBM SPSS : and sex and drugs and rock 'n' roll*, London, SAGE.

FRANKLIN & ANDREWS 2011. *Hutchins UK Building Blackbook: The Cost and Carbon Guide : Hutchins' 2011: Small and Major Works*, Croydon, Franklin & Andrews.
HACKER, J. N., DE SAULLES, T. P., MINSON, A. J. & HOLMES, M. J. 2008. Embodied and operational carbon dioxide emissions from housing: A case study on the effects of thermal mass and climate change. *Energy and Buildings,* 40(3)**,** 375-384.
HALCROW YOLLES 2010a. Embodied Carbon Sustainable Offices: a supplementary report. UK: South West Regional Development Agency.
HALCROW YOLLES 2010b. Sustainable offices – embodied carbon. UK: South West Regional Development Agency.
HAMMOND, G. & JONES, C. 2011. A BSRIA guide Embodied Carbon The Inventory of Carbon and Energy (ICE). UK: BSRIA.
KARANCI, H. 2010. *A comparative study of regression analysis, neural networks and case –based reasoning for early range cost estimation of mass housing projects.* Master of Science, Middle East Technical University.
KARSHENAS, S. 1984. Predesign cost estimating method for multistory buildings. *Journal of Construction Engineering and Management,* 110(1)**,** 79-86.
KIM, G.-H., AN, S.-H. & KANG, K.-I. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and environment,* 39(10)**,** 1235-1242.
KNEIFEL, J. 2010. Life-cycle carbon and cost analysis of energy efficiency measures in new commercial buildings, *Energy and Buildings*, 42(3), 333-340.
KOHAVI, R. 1995. A study of cross-validation and bootstrap for accuracy estiamtion and model selection. *In International Joint Conference on Artifical Intelligence San Francisco, CA, 1137–1143.*
KOUSKOULAS, V. & KOEHN, E. 2005. Predesign cost estimating function for buildings. *In:* SKITMORE, M. & MARSTON, V. (eds.) *Cost Modelling.* London: E & FN Spon.
LUO, Z., YANG, L. & LIU, J. 2015. Embodied Carbon Emissions of Office Building: A Case Study of China's 78 Office Buildings [online]. *Building and Environment*.
MILES, J. & SHEVLIN, M. 2001. *Applying regression & correlation: a guide for students and researchers,* London: SAGE.
MONAHAN, J. & POWELL, J. C. 2011. An embodied carbon and energy analysis of modern methods of construction in housing: A case study using a lifecycle assessment framework. *Energy and Buildings,* 43(1)**,** 179-188.
OLIVIER, J. G. J., JANSSENS-MAENHOUT, G., MUNTEAN, M. & PETERS, J. A. H. W. 2016. Trends in global CO2 emissions. The Hague: PBL Netherlands Environmental Assessment Agency; Ispra: European Commission, Joint Research Centre.
PERERA, S. & VICTORIA, M. 2017. The role of carbon in sustainable development. *In:* LOMBARDI, P., SHEN, G. Q. & BRANDON, P. S. (eds.) *Future challenges for sustainable development within the built environment.* UK: Wiley's Publication.
PEURIFOY, R. L. & OBERLENDER, G. D. 2002. Estimating construction costs. 5th ed. Boston, Mass, London: McGraw-Hill.
RIBA. 2013. *RIBA plan of work 2013* [Online]. Available: http://www.ribaplanofwork.com/ [Accessed 20th October 2014].
RICS 2009. RICS oceania - expert witness registration scheme. UK: RICS.
RICS 2014. Methodology to calculate embodied carbon. 1 ed. UK: RICS.
RICS 2016. Building Cost Information Services (BCIS). RICS.
SANSOM, M. & POPE, R. J. 2012. A comparative embodied carbon assessment of commercial buildings. *The Structural Engineer,* October**,** 38-49.

THE GREEN CONSTRUCTION BOARD 2013. Low carbon routemap for the uk built environment. London: The Green Construction Board, WRAP, The Climate Centre and ARUP.
UK-GBC 2014a. Building zero carbon – the case for action examining the case for action on zero carbon non domestic buildings. UK: UK-GBC.
UK-GBC. 2014b. *Key statistics: construction industry and carbon emissions* [Online]. Available: http://www.ukgbc.org/resources/additional/key-statistics-construction-industry-and-carbon-emissions [Accessed 27 February 2017].
VICTORIA, M., PERERA, S. & DAVIES, A. 2015a. Developing an early design stage embodied carbon prediction model: a case study *In:* RAIDÉN, A. B. & E., A.-N. (eds.) *31st Annual ARCOM Conference, 7-9 September 2015.* Lincoln, UK: Association of Researchers in Construction Management.
VICTORIA, M. F., PERERA, S., ZHOU, L. & DAVIES, A. 2015b. Estimating embodied carbon: a dual currency approach. *Sustainable Buildings and Structures.* CRC Press.
WRAP & UK-GBC. 2014. *Embodied carbon database* [Online]. WRAP. Available: http://www.wrap.org.uk/content/embodied-carbon-database [Accessed 27 February 2017].
WU, H., YUAN, Z., ZHANG, L. & BI, J. 2012. Life cycle energy consumption and CO2 emission of an office building in China. *The International Journal of Life Cycle Assessment,* 17**,** 105-118.
YEO, Z., NG, R. & SONG, B. 2016. Technique for quantification of embodied carbon footprint of construction projects using probabilistic emission factor estimators, *Journal of Cleaner Production*, 119, 135-151.
YOHANIS, Y. G. & NORTON, B. 2002. Life-cycle operational and embodied energy for a generic single-storey office building in the UK. *Energy*, 27(1)**,** 77-92.

Figure 1: Embodied carbon values of different types of buildings from the literature

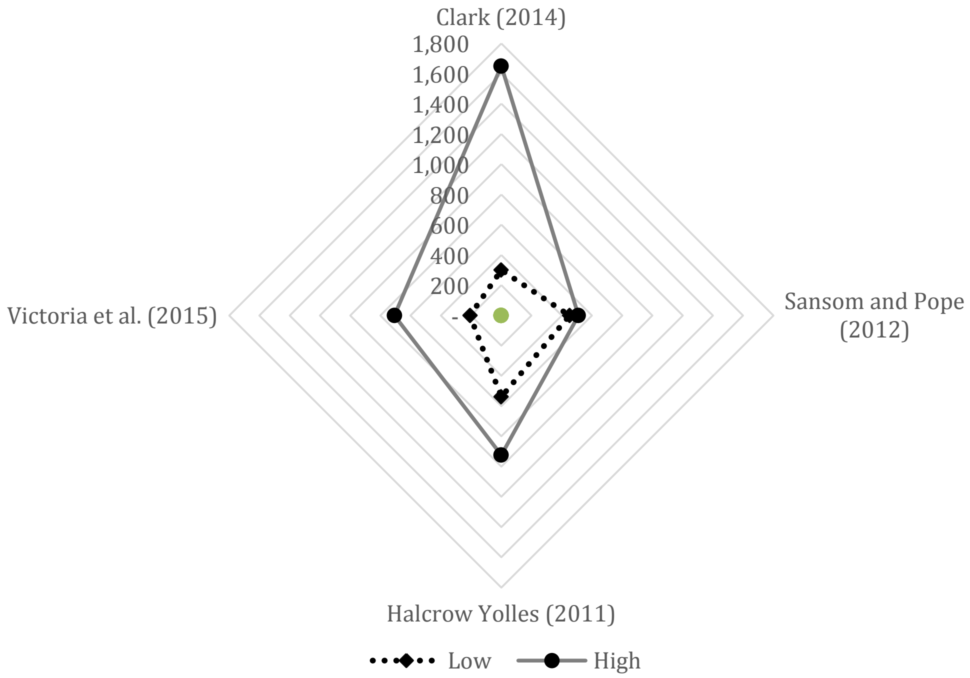


Figure 2: Embodied carbon studies on office buildings

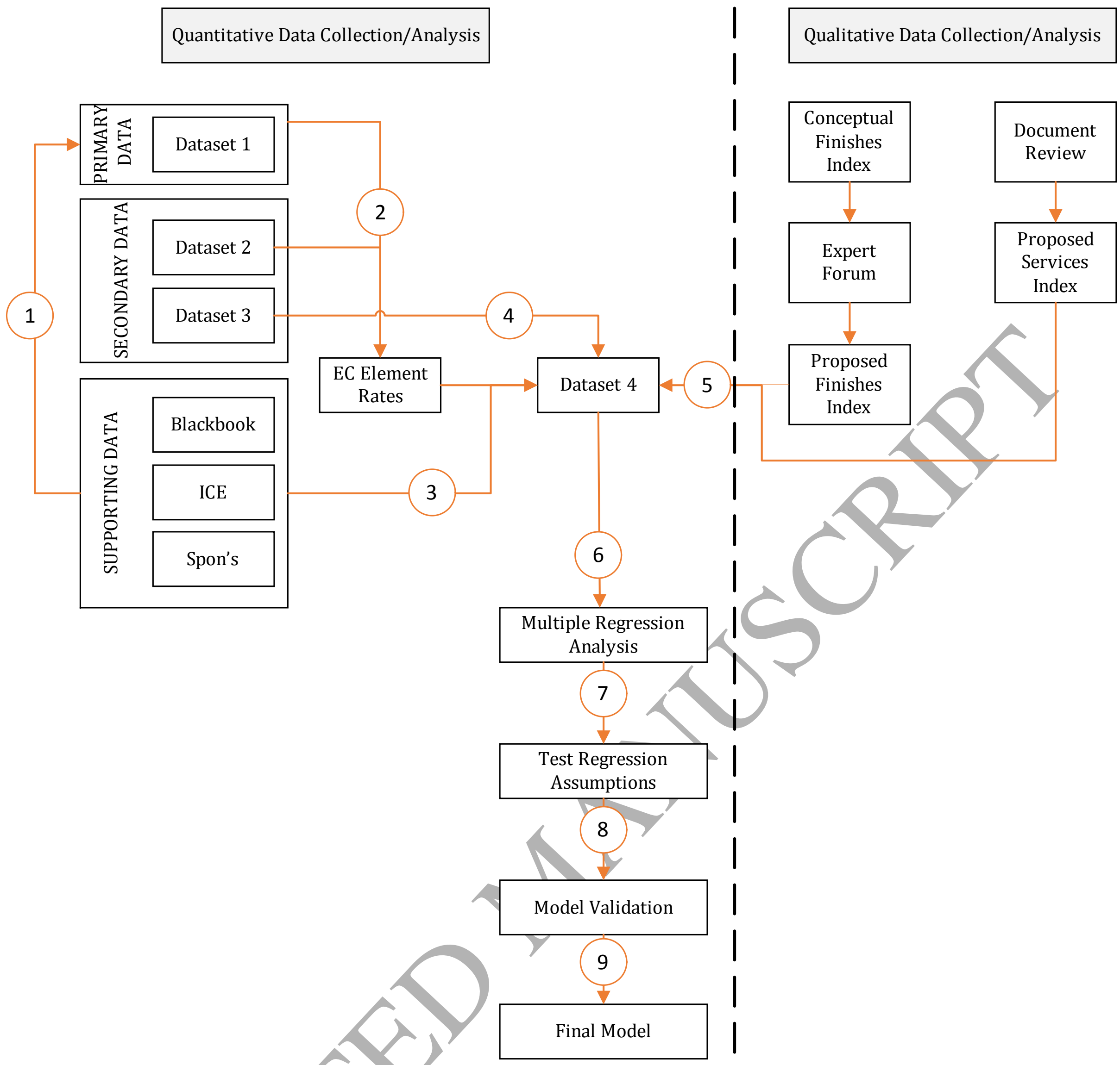


Figure 3: Overview of the methodology

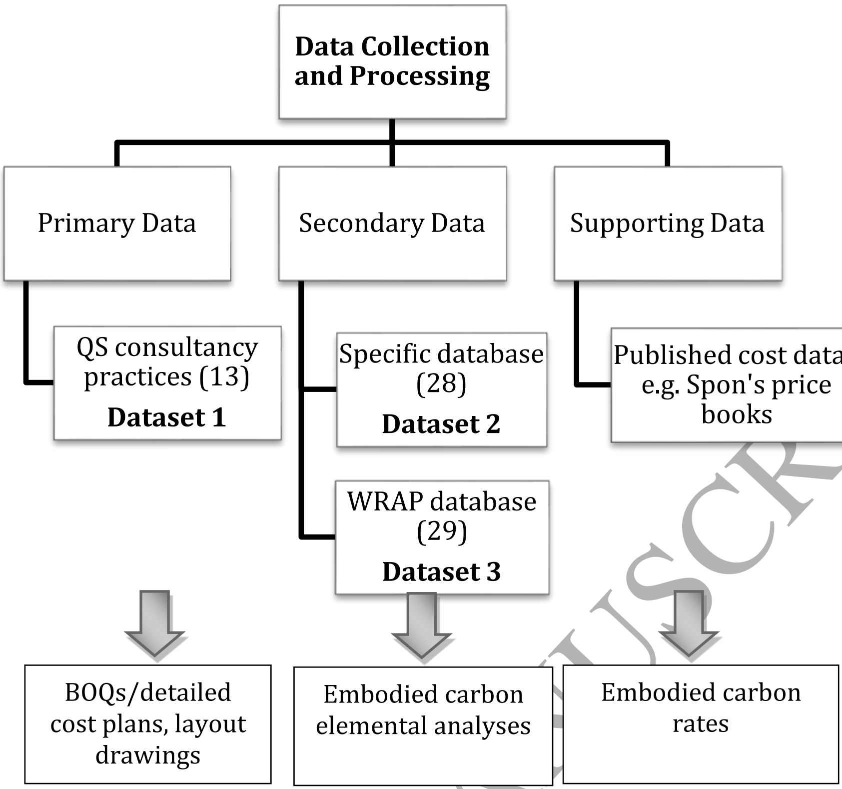


Figure 4: Types and sources of data obtained (numbers within the brackets denote the sample size)

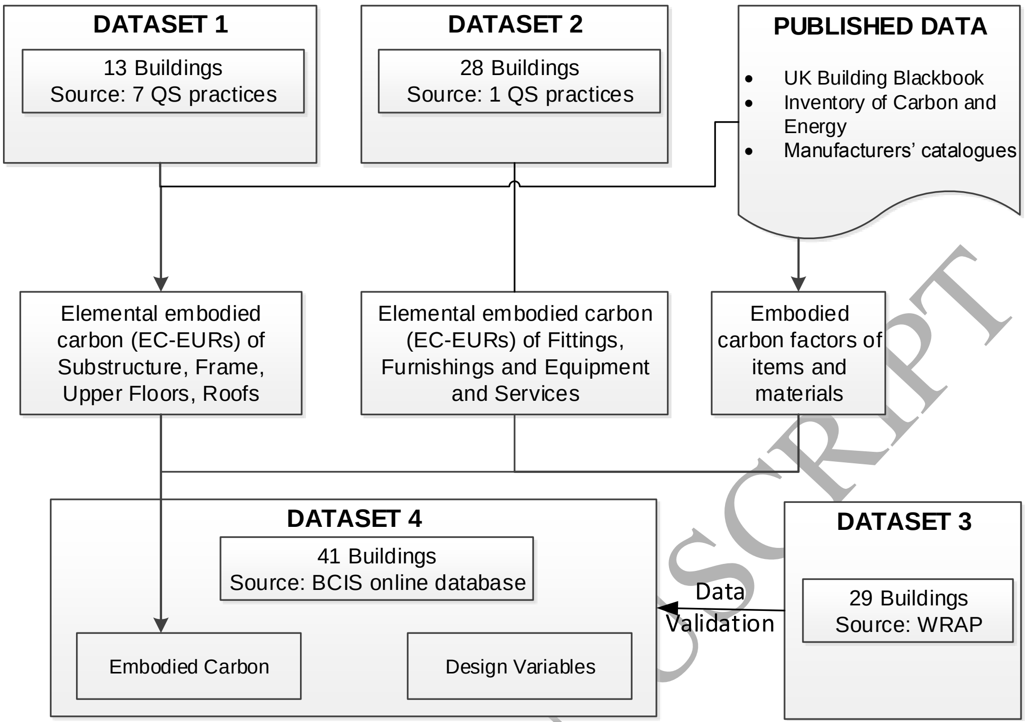


Figure 5: Development and validation of Dataset 4

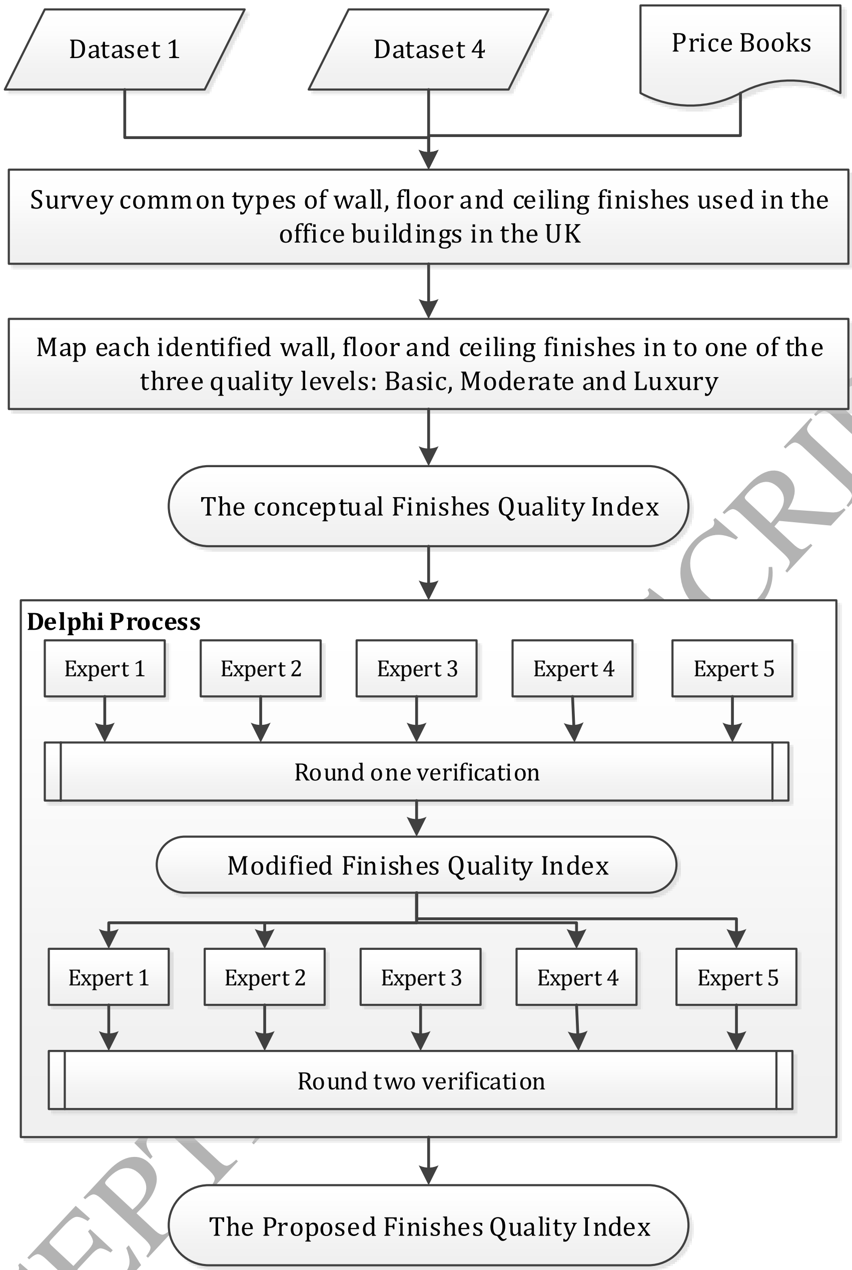


Figure 6: The development process of the finishes quality index

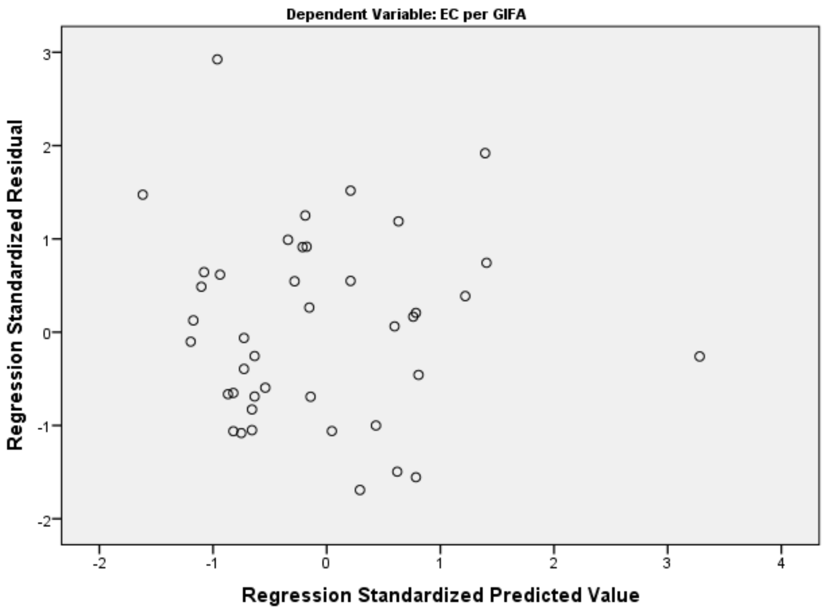


Figure 7: Scatterplot of standardised predicted value vs. standardised residuals of regression

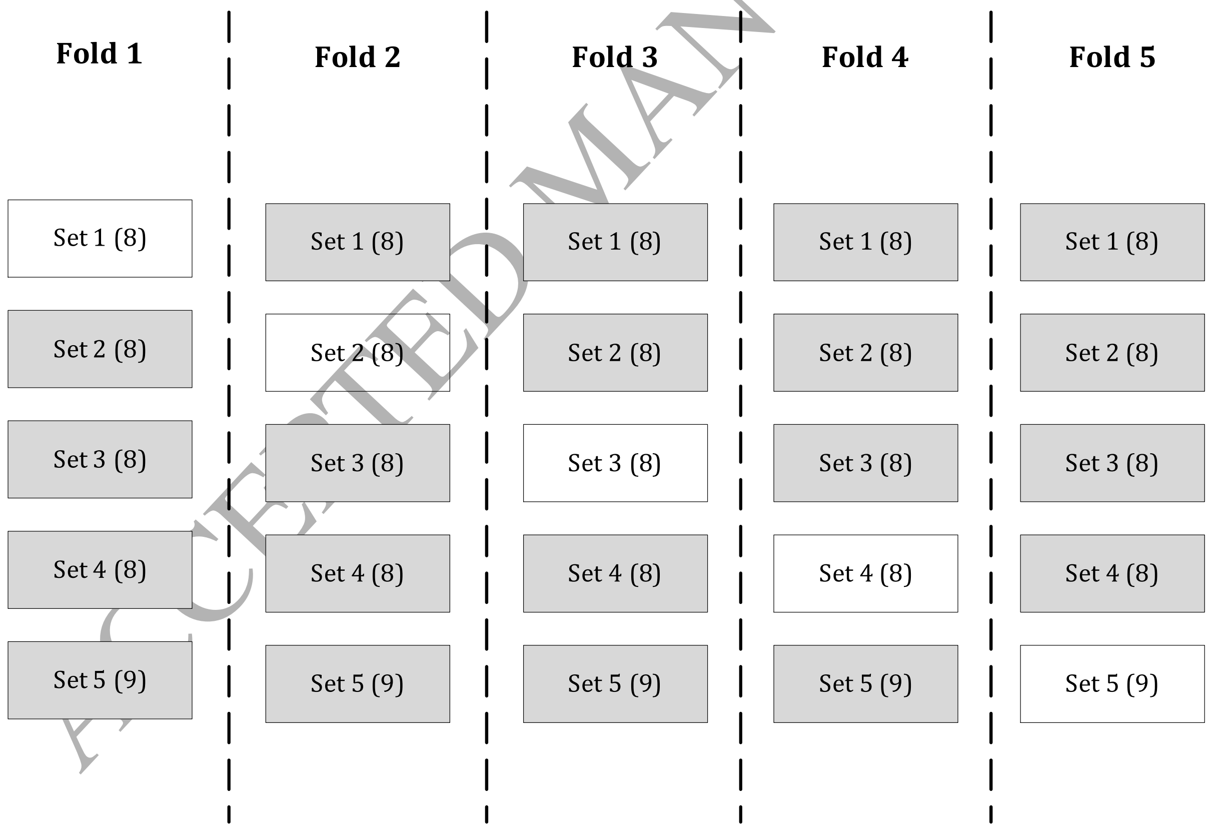


Figure 8: Five-fold cross validation data splitting of Dataset 4

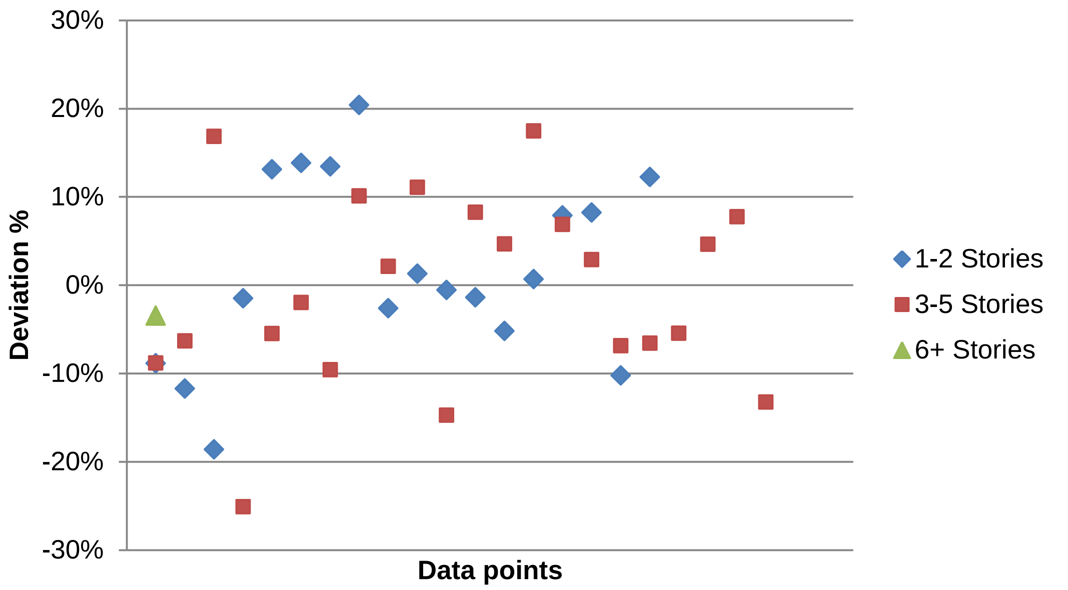


Figure 9: The model predictions at different storey clusters

Table1: Summary of notations used

Notations	Definition
$EC_{m/i}$	Total embodied carbon of a particular material or an item in the building considered
$Q_{m/i}$	Total quantity of the respective material or item in the building considered
$ECF_{m/i}$	Embodied carbon factor of the respective material or item (i.e. $kgCO_2$/kg of the material OR $kgCO_2$/unit of the item – $kgCO_2/m^3$ of concrete etc.)
$EC-EUR_i$	Embodied Carbon Element Unit Rate of element '*i*'
EC_i	Embodied carbon of element '*i*'
EUQ_i	Element Unit Quantity of element '*i*'
$FI_{Building}$	Overall finishes quality of the building
A_i	Area of wall/floor/ceiling finishes as a percentage of the total finished area
$FI_{W,F,C}$	Overall wall/floor/ceiling finishes index
a_i	Area of basic/moderate/luxury finishes as a percentage of the total wall/floor/ceiling finishes area
I_i	Respective index (Basic – 1, Moderate – 2, Luxury – 3)
$\hat{y}$	Estimated Embodied Carbon per GIFA of the building
a_0	Regression constant
a_1	Regression coefficient of $x_{W:F}$
$x_{W:F}$	Wall to Floor ratio of the building
a_2	Regression coefficient of x_{ASH}
x_{ASH}	Average Storey Height of the building
a_3	Regression coefficient of x_{BH}
x_{BH}	Building Height
a_{FI}	Finishes Index of the building
a_{SI}	Services Index of the building
d	Durbin-Watson test statistics
e_i	Residual of the i^{th} observation
n	Number of observations
$d_{L,\alpha}$	Lower critical value of the Durbin Watson test statistics for a given significance level (α)
$d_{U,\alpha}$	Upper critical value of the Durbin Watson test statistics for a given α
CV	Coefficient of Variation
$\overline{e}$	mean of the residuals
$\overline{y}$	mean of the observed values (actual embodied carbon values)

Table 2: A review of embodied carbon estimating practices adopted in past studies

Study	RIBA 2013 Stage	System boundary	Source of EC data	Estimating Technique
Halcrow Yolles (2010)	4 – Technical design	Cradle-to-Gate	The UK Building Blackbook	Bottom-up approach
Victoria et al. (2015)	4 – Technical design	Cradle-to-Gate	The UK Building Blackbook	Bottom-up approach
Sansom and Pope (2012)	4 – Technical design	Cradle-to-Grave (excl. recurring emissions)	GaBi database	CLEAR life cycle assessment model/ bottom-up approach
Monahan and Powell	4 – Technical design	Cradle-to-Site	ICE, ecoinvent, published government sources, US life cycle inventory	Simapro software/ bottom-up approach
Hacker	4 – Technical design	Cradle-to-Grave	Published data from Institution of Structural Engineers	Bottom-up approach
Sturgis and Roberts (2010)	4 – Technical design	Cradle-to-Grave	ICE, conversions factors from Department of Environment, Food and Rural Affairs (DEFRA), BCIS lifespan data	Bottom-up approach
RICS (2014)	4 – Technical design	Cradle-to-Gate	ICE, SimaPro, GaBi	Bottom-up approach
	5 – Construction 6 – Handover and closeout	Gate-to-Construction	DEFRA Greenhouse Gas Conversion Factor Repository, GHG Protocol calculation tools	Bottom-up approach
	7 – In Use	Construction-to-Grave	BCIS Life Expectancy of Building Components (BCIS 2006) + product stage sources	Bottom-up approach
Yeo et. al (2016)	3 – Developed Design 4 – Technical design	Cradle-to-Gate	ICE, ecoinvent, World Steel Association, Franklin USA etc.	Probabilistic method
Construction Carbon Calculator	2 – Concept Design	Cradle-to-Construction	Web-based resources of embodied carbon intensity ratios of different building materials.	Parametric model (methodology is not transparent)
Steel Construction Embodied Carbon Tool (structure only)	3 – Developed Design	Cradle-to- Grave (excl. recurring emissions)	Environmental Product Declaration (EPD) published by the European steel industry	'Auto generated mode' estimate structural material quantities using algorithms. 'Manual input' mode allows to enter the actual material quantities
Embodied CO_2 Estimator	3 – Developed Design	Cradle-to-Construction (excluding transport)		Not explicit though it appears to be underpinned by some form of algorithm
Carbon calculator for construction projects	4 – Technical design	Cradle-to- Grave		Bottom-up approach

Table 3: Mapping each dataset against the data requirement

Required data	Measurement scale	Dataset 1	Dataset 2	Dataset 3
1. Measurement of quantities/ element unit quantities of the buildings	Ratio	Yes (some elements are not measured)	No	No
2. Specification of the buildings	Nominal	Yes	No	No
3. Design variables of the buildings – quantitative	Ratio	Yes	Yes (Only GIFA & no. of storeys)	Yes (Only GIFA & no. of storeys)
4. Design variables of the buildings – qualitative	Ordinal	Yes	No	No
5. Embodied carbon	Ratio	Yes (Excludes Fittings & Services)	Yes	Yes (Excludes Fittings & Services)

Table 4: Element groups as prescribed in WRAP dataset (Dataset 3)

Element Category	Elements included (as per the NRM)
Substructure	Substructure – foundations, basements and ground floor
Superstructure Structural	Frame, Upper Floors and Roof
Superstructure Non-Structural	Internal Walls and Partitions and Internal Doors
Internal Finishes	Wall Finishes, Floor Finishes and Ceiling Finishes

Table 5: Group statistics for individual element categories

Element Group	Group	Sample size	Mean	Std. Deviation	Kolmogorov-Simirnov		Kolmogorov-Simirnov after log transformation	
					Statistic	Sig.	Statistic	Sig.
Substructure	Dataset 3	29	146.08	74.13	0.081	0.200	0.132	0.200
	Dataset 4	41	161.158	57.53	0.160	0.010	0.117	0.170
Superstructure Structural	Dataset 3	29	363.84	116.01	0.136	0.183	N/A	
	Dataset 4	41	219.45	63.80	0.070	0.200		
Superstructure Non-Structural	Dataset 3	29	34.67	49.77	0.290	0.000	0.090	0.200
	Dataset 4	41	25.40	33.76	0.300	0.000	0.134	0.060
Internal Finishes	Dataset 3	29	55.68	36.87	0.171	0.030	0.230	0.000
	Dataset 4	41	54.64	16.06	0.178	0.002	0.122	0.132

Table 6: t-Test statistics of the two samples – Dataset 3 and Dataset 4

Element Category		Levene's Test for Equality of Variances		t-Test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Log of Substructure	Equal variances assumed	6.496	0.013	-1.666	68	.100
	Equal variances not assumed			-1.525	40.954	.135
Superstructure Structural	Equal variances assumed	5.030	.028	6.680	68	.000
	Equal variances not assumed			6.083	39.979	.000
Log of Superstructure Non-Structural	Equal variances assumed	1.049	.309	.803	68	.425
	Equal variances not assumed			.779	53.467	.439

Table 7: The proposed finishes quality index for the study

	Basic (Index – 1)	**Moderate (Index – 2)**	**Luxury (Index – 3)**
Wall Finishes	Paint to fair face, Cement sand plaster, Emulsion/eggshell, Lining paper, Vinyl paper, Basic ceramic tiles	Thistle plaster, Carlite plaster, Moisture resistant paint, Plasterboard, Plywood wall panels and treatment, Wallboards Softwood boarding and treatment, Hardboard Chipboard, Veneered MD panels, Moderate ceramic tiles, Moderate porcelain tiles, Chinese marble	Mosaic tiles, Luxury ceramic tiles, Luxury porcelain tiles, Heavily embossed wallpapers, Natural granite , European marble, Composite aluminium, Glass
Floor Finishes	Concrete hardener, Regular floor paint, Cement sand, Latex screed, Mastic asphalt floor, Linoleum sheet, Linoleum tiles, Basic vinyl sheet, Basic vinyl tiles, Basic carpet tiles, Cement tiles, Basic ceramic tiles, Medium duty carpet	Granolithic, Epoxy floor, Rubber floor tiles, Marmoleum, Moderate vinyl sheet, Moderate vinyl tiles, Cork tiles, Moderate carpet tiles, Moderate ceramic tiles, Moderate porcelain tiles, Clay tiles, Quarry tiles, Heavy duty carpet, Terrazzo, Chinese marble, Metal access floors, Veneered laminated floor Redwood floor	Mosaic tiles, Slate tiles, Luxury ceramic tiles, Luxury porcelain tiles, Woodblock floor (Oak etc.), Woodstrip floor (Oak etc.), Parquet floor, Natural granite, European marble
Ceiling Finishes	Sealer, Skim coat, Cement sand plaster, Emulsion/eggshell, Lining Paper	Thistle plaster, Carlite plaster, Moisture resistant paint, Plasterboard, Metal frame plasterboard ceilings, Plasterboard acoustic ceilings, Moisture resistant ceilings, Metal suspended ceilings	Timber boarded ceilings, Moisture resistant ceilings with high sound proofing Coffered ceilings

Table 8: The proposed service quality index

Services Quality Index	
Level 1 - Non air-conditioned buildings (Essential building services)	
1.1	Without lift
1.2	With lift
Level 2 - Air-conditioned buildings (Level 1 + A/C)	
2.1	Without lift
2.2	With lift
Level 3 - Non air-conditioned automated buildings (Level 1 + BMS)	
3.1	Without lift
3.2	With lift
Level 4 - Air-conditioned automated buildings (Level 2 + BMS)	
4.1	Without lift
4.2	With lift

Table 9: Descriptive statistics of variables

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
						Statistic	Std. Error
GIFA	41	212	14652	3642.07	3329.495	1.535	.369
Building Height	41	2.8	25.2	9.50	3.828	1.756	.369
Wall to Floor Ratio	41	.24	1.50	.71	.243	.926	.369
Circulation Ratio	33	.09	.46	.24	.092	.477	.409
Embodied Carbon per GIFA	41	551	916	680.44	95.581	.696	.369

Table 10: Correlation matrix

		Building Height	Wall to Floor Ratio	Circulation Ratio	EC per GIFA
Building Height	Pearson Correlation	1	.206	.113	.306
	Sig. (2-tailed)		.195	.531	.052
	N	41	41	33	41
Wall to Floor Ratio	Pearson Correlation	.206	1	.304	.523**
	Sig. (2-tailed)	.195		.086	.000
	N	41	41	33	41
Circulation Ratio	Pearson Correlation	.113	.304	1	.360*
	Sig. (2-tailed)	.531	.086		.039
	N	33	33	33	33
EC per GIFA	Pearson Correlation	.306	.523**	.360*	1
	Sig. (2-tailed)	.052	.000	.039	
	N	41	41	33	41

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 11: Summary of the models produced using the backward regression analysis

Model	R^2	Adjusted R^2	F Statistics	Sig.	Std. Error of the Estimate	Independent Variables
1	.559	.457	5.484	.001	72.011	Building height, wall to floor ratio, circulation ratio, no. of basements, finishes index, services index
2	.557	.475	6.794	.000	70.781	Wall to floor ratio, circulation ratio, no. of basements, finishes index, services index
3	.552	.488	8.620	.000	69.922	Wall to floor ratio, no. of basements, finishes index, services index
4	.542	.495	11.440	.000	69.456	Wall to floor ratio, no. of basements, finishes index
5	.513	.481	15.828	.000	70.386	Wall to floor ratio, no. of basements

Table 12: Summary of five-fold cross validation outcomes

Fold	Model	Adj. R^2	Predictors	CV	
				Training Set	Test set
Fold 1	Model 1	0.316	$867 + 136x_{W:F} + 52x_B - 154x_{FI}$	10.74%	11.12%
Fold 2	Model 2	0.476	$504 + 201x_{W:F} + 62x_B$	10.39%	10.75%
Fold 3	Model 3	0.490	$504 + 181x_{W:F} + 59x_B$	9.6%	14.37%
Fold 4	Model 4	0.398	$565 + 122x_{W:F} + 70x_B$	10.36%	12.07%
Fold 5	Model 5	0.423	$559 + 135x_{W:F} + 69x_B$	10.42%	11.25%
Average				10.30%	11.91%