West Africa Built Environment Research (WABER) Conference
WABER 2021 Conference, 9-11 August 2021, Accra, Ghana
Proceedings Edited by Laryea, S. and Essah, E.
ISBN 978-0-620-95367-2

CONCEPTUAL FRAMEWORK FOR WHOLE-LIFE COST DATA TRANSFORMATION AND MODEL SELECTION IN THE BUILDING SECTOR

Ibrahim, A. M.[1], Bala, K.[2], Ibrahim, A. D.[3] and Zubairu, I. K.[4]
[1,2,4]*Department of Building, Faculty of Environmental Design, Ahmadu Bello University, Zaria, Nigeria.*
[3]*Department of Quantity Surveying, Faculty of Environmental Design, Ahmadu Bello University, Zaria, Nigeria*

Data for Whole Life Costing (WLC) is obtained from multiple sources in different formats. This inhibits data sharing between built-environment costs professionals thereby adding to data related problems often associated with WLC. This paper presents a proposed framework for mapping WLC data to models, as well as transforming data into a variety of formats. Framework components, variables and processes were identified from literature and synthesized into a conceptual framework. Also, a matrix of transformation logic was produced to compliment the model selection process. Seven data related scenarios and possible decisions/ courses of actions were derived from the framework. Additionally, seven useful data types and fourteen data formats were identified. Of the fourteen formats random statistical data, probability density functions and fuzzy membership functions were found to be the most important in terms of transformability ranking. The framework provides strategic decision guide for cost professionals during the preparation of WLC data and the selection of a suitable model, while the transformation processes provides a procedural guide for converting data into different formats. The findings set the stage for the development of transformation and integration algorithms to support computer coding, and the design and implementation of a database that would facilitate the storage and retrieval of data in multiple formats.

Keywords: data integration, data sharing, data transformation, whole life cost data, whole life cost models

INTRODUCTION

Sustainable development agenda requires building projects to be economically efficient so that, they are free of under or over investment over their life cycle. An indispensable tool that helps in the evaluation of economic sustainability of a building is the Whole Life Costing (WLC) technique (Boussabaine & Kirkham 2004; Ellingham & Fawcett 2006). As a technique, WLC is used for the assessment of the

[1] alibramak@gmail.com
[2] balakabir@yahoo.com
[3] adibrahim2@yahoo.com
[4] ibrakhazu@gmail.com

cradle-to-grave cost of a building or its part, component or system. It represents the present value of the sum of the initial, operating, maintenance, and terminal costs of a built asset. It is essentially used as a decision-making tool to facilitate the selection of the best option (design, component, building, system etc) from a host of competing alternatives; for design optimisation; and for budgetary purposes (Flanagan & Jewell 2005; Hoar 2007). A typical WLC requires cost, physical, economic, and performance data. These data come from diverse sources in different forms and varying characteristics like tangibility, certainty and availability (Kishk et al. 2003). The application of the technique in especially the developing countries, despite its usefulness, is retarded by a number of challenges. Higham, Fortune, and James (2015), and Opawole et al. (2020), identified these challenges to include; data, client, process, politics, government, practice, and tools related.

The data related problem centres on data scarcity, difficulty in data sharing, and inconsistent databases. Firstly, the dearth of reliable historical data makes WLC expensive to conduct due to cost of data collection which reduces confidence in the reliability of the results (Al-Hajj et al. 2001). Secondly, data for WLC is elicited from different sources and in different structure, which necessitates data adjustment in order to bring it to a common basis in terms of type, structure, time, and location. Such adjustment is imperative to data sharing between different models, if model-data mismatch is to be avoided (Saridaki, Psarra, & Haugbølle 2019). Data adjustment or transformation is a complex process and unless it is simplified, data elicitation and processing will remain tedious thereby aggravating the data scarcity problem. Thirdly, existing databases are inconsistent and inaccurate. Al-Hajj et al. (2001) described them as 'constructed 'rather than 'historical-based 'with data background information often not recorded making them unreliable.

Kishk et al. (2003) underscore the importance of data transformation during the selection of a modelling technique. They proposed an integrated framework for WLC that facilitates the selection of an appropriate modelling technique on the basis of data tangibility, availability and certainty. The framework however considers only the two extremes of data availability (available data, and no data), eliminating possibilities of having incomplete data as suggested by Ilg et al. (2017).

The aim of this paper is to develop a WLC framework that facilitates data transformation, as well as the identification and selection of an appropriate WLC model and data format. It is strongly believed that built-environment cost professionals will find the framework a useful guide for data/model matching, data transformation, and sharing as well as providing the basis for the development of a robust WLC database.

LITERATURE REVIEW

Purpose of WLC

Whole Life Costing (WLC) can be described as a technique that facilitates the systematic computation of the total cost of an asset or its part, from concept to its end-of-life (Hoar 2007; Norman 2007). It is essentially used to aid decision-making when choosing between competing alternatives (building investment options,

design alternatives, systems, or components); optimisation of designs, performance, system, or components; and forecast of costs for budgetary purpose (Kishk et al. 2003; Hoar 2007).

Data requirement for WLC

Data for WLC can be both quantitative and qualitative. Quantitative data is numerical in nature and represents cost data (like initial, operating, maintenance repair, and terminal cost), physical data associated with dimension of a building (e.g. gross floor area, floor to ceiling height, external perimeter etc), economic data (like interest rate, inflation rate, discount rate etc), and performance data (like maintenance cycles, life expectancy, occupancy time, electricity consumption, water consumption, gas consumption etc). These data are useful in WLC using quantitative techniques like Net Present Value (NPV) (Kishk et al. 2003; Flanagan & Jewell 2005). Qualitative data on the other hand is subjective and often derived from expert judgement/opinion (e.g. condition, quality, aesthetics, comfort and maintenance ratings, intangible costs and benefits). These kinds of data are used in qualitative WLC using weighted evaluation techniques (Norman 2007). Dell'Isola and Kirk (2003) proposed a pair-wise weighted evaluation technique suitable for analysing qualitative WLC data. Kishk (2002) and Kishk, Al-Hajj, and Pollock (2002) modified the technique using Fuzzy Set Theory (FST). Norman (2007) noted that, it is imperative to have a blend of both quantitative and qualitative approach so that computed WLC results can be moderated with professional judgement and experience.

Data related challenges

The dearth of historical data has been reported by many authors (Flanagan & Jewell 2005; Higham, Fortune, & James 2015) and remains the most important challenge to WLC. In an attempt to mitigate the scarcity problem, Kishk and Al-Hajj (2000) developed an innovative FST based approach that utilizes expert opinion and judgement. Another data related problem is the lack of uniformity in the source, type and structure of WLC data. This means that before data can be used in a different setting, substantial adjustments (often complex and costly) must be made to normalise and express it on a common scale. In order to simplify the adjustment process, there is the need to automate the process so that data can easily be converted and shared for use in different environment.

Another data related problem resides with the limitations of existing databases. These databases which are supposedly designed to provide a system that simplifies WLC data storage and retrieval are rather inconsistent. Al-Hajj et al. (2001) argued that the databases are based on "expert opinion", trades publication data, and manufacturers 'data, and that, background information and measures of uncertainty are often lost during data storage as the context information on the data is often not recorded. This is particularly a problem as data relating to a proposed project is often generated from multiple sources.

Quantitative WLC data formats

Quantitative data used in WLC comes in different formats depending on availability and nature of uncertainty. Crisp, single or point data is a discrete and precise value, free of variability and thus used in deterministic WLC (Boussabaine & Kirkham 2004; Ashworth & Parera 2016). Where the nature of variability in the data is not

clear, interval data can be used to represent the uncertainty. Range of values are used to mark the lower and upper bounds of the data. A variable represented by interval data can take any value within the bound (Bruns, Paredis, & Ferson 2006). WLC can then be done deterministically using interval arithmetic. Another format is the three-point data that marks the lower and upper bounds, as well as the best estimate of the data (Boussabaine & Kirkham 2004). Sensitivity analysis can be performed within the data bound to identify the most sensitive variables (Flanagan et al. 1987). A variant of the three-point data introduced by Dell'Isola and Kirk (2003) for use in Confidence Index (CI) approach to WLC, is conditional, and require the data to meet dual criteria. First the high and low 90% must be obtained from same source as the best estimate and should represent knowledgeable estimate rather than guess work. Second, the difference between the present values of the best estimates and the present value of the lower and upper 90% bounds must be within 25% of each other. The approach is designed around probability theory and assumes uncertainty in WLC data to be normally distributed, and the lower and upper 90% bounds to correspond to the true 90% mark on a normal probability distribution.

Where sufficient sample of historical data is available, objective frequency distributions can be generated and a suitable representative probability distribution identified (Kishk & Al-Hajj 2001). Bossabaine and Kirkham (2004) identified some popular PDF as normal, lognormal, beta, Weibull, Pareto, exponential, and gamma. The PDF can then be used in simulation models. Alternatively, such data can form the basis for developing regression models (Ashworth & Parera 2016). Statistical data may however turn out to be non-random which makes it unsuitable for stochastic modelling. In such a situation, other modelling techniques like Artificial Neural Network (ANN), Genetic Algorithms (GA) and Fuzzy Set Theory (FST) or their combination may be deployed. Ross (2004) demonstrated how fuzzy sets can be generated through intuition, inference, rank ordering, ANN, GA, and inductive reasoning.

Sample size in statistics is an important determinant of the kind of analysis to apply. Spiegel and Stephens (1999) noted that, small sampling theory requires a sample size of less than thirty for its application. In fact, Stephen and Berenson (2005) showed that, the central limit theorem holds where the sample size is at least thirty for a population that is not normally distributed.

Where historical data is not sufficient to form a definitive distribution, as a result of say missing data, heuristics may be used to augment the data (Goh et al. 2010) based on expert judgement. Data augmentation techniques like Convolution Neural Networks (CNN) (Zheng et al. 2020), Bayesian data augmentation approach (Tran et al. 2017), and Markov-chain Monte Carlo (van Dyk & Meng 2001) have been successfully used in the fields of medicine, imagery, music amongst others. There usability for WLC needs to be explored.

Subjective probabilities are used where WLC data is not available. These probabilities are systematically collected from experts and a subjective PDF formulated. The PDF can take the form of uniform, triangular, and trapezoidal density functions (Seeley 1998). A procedure for the development of subjective probabilities has been recommended by Galway (2007). Wang, Chang and El-

Sheikh (2012) have demonstrated the development and use of subjective probabilities during simulation of WLC.

Another form of data used in WLC is linguistic variables. These are expert statements that are either ambiguous, vague or imprecise. Such statements are best expressed using fuzzy numbers (Zimmermann 2001; Ross 2004). Fuzzy numbers can take different shapes like rectangular, triangular, trapezoidal and Gaussian. Kisk and Al-Hajj (2000) have demonstrated the application of Fuzzy Set Theory in WLC of building projects.

WLC Models

A model is a prototype of a real object or its characteristics, procedure, or system (Anderson, Sweeney & Williams 2004; Fellows & Liu 2007). The goal is for the prototype to closely represent the real object. Models come in different forms depending on how they mimic reality. A model that is a replica of a real object is called iconic model (Kirkham 2015). Analog models are physical in form but are not exact replica of the object they represent. Symbolic or mathematical models represent a scenario with a system of symbols, logical or mathematical expressions (Kirkham 2015). Another classification is based on the transparency of model's internal structure. In this class models can be either black-box or white-box. The internal structure of a black-box model is opaque, that is not completely known to the modeller. Such models use some form of heuristics to generate output based on some given input data. Examples are Artificial Neural Networks (ANN), and Genetic Algorithms (GA). The internal structure of a white-box model is transparent in the sense that, it is completely known to the modeler. Symbolic models are transparent as their structure is known to the modeller (Fellows & Liu 2007).

Two broad classes of cost models have been identified by Skitmore and Marston (1999). These are cost-product and cost-process models also referred to as 'Designers 'cost models 'and 'Constructors 'or 'production cost models 'respectively by Ashworth and Parera (2016). The cost-product model relies on information from completed projects to model cost of proposed developments. Cost-process models on the other hand mimic construction process based on operations or activities often at a detailed level, to model cost of proposed projects (Kirkham 2015). Another important classification of cost model is based on the way uncertainty is treated. Deterministic models assume perfect information. They take precise cost data as input and produce output that is interpreted as certain (Ashworth & Parera 2016). The second group is stochastic model which recognise reality to be full of uncertainty and thus model cost as a random parameter based on probability theory (Bowen, Wolvart & Tailor 1987; Ashworth & Parera 2016). The third group expresses uncertainty as non-random and model cost in terms of fuzzy sets or heuristics (Boussabaine and Kirkham 2004).

The purpose of WLC models is to ease process complexity and improve efficiency in achieving the core objectives of WLC. Boussabaine and Kirkham (2004) identified three WLC modelling techniques. They include deterministic, stochastic and FST based techniques. Kim, An, and Kang (2004) also described how regression analysis can be used to model the relationship between variables in WLC.

Uncertainty and risk in WLC data

The desire to predict decisions, performances and costs that will occur many years into the future, coupled with the dearth of reliable historical data makes WLC prone to uncertainty (El-Haram et al. 2007). Zimmerman (2001) described uncertainty as the lack of quantitative and qualitative information needed to adequately prescribe or predict both deterministically and numerically the characteristics of a system. He further identified the causes of uncertainty as lack of information, overabundance of information, conflicting evidence, ambiguity, measurement error, and belief. Huijbregts et al. (2003) identified three types of uncertainty in WLC: parameter (data related uncertainty), model (uncertainty in mathematical relationship) and scenario (uncertainty in normative choices).

Parametric uncertainty is data related, and occurs due to lack of information, over-abundance of information, measurement error, ambiguity, and unrepresentative data (Huijbregts 1998; Zimmerman 2001). Goh et al. (2010), further classified parametric uncertainty into aleatory and epistemic. They described aleatory uncertainty as intrinsic variability in data that does not completely vanish with additional information. Epistemic parametric uncertainty occurs primarily due to lack of knowledge on the true value of a parameter and can be reduced with additional information. It is more difficult to assess and thus require new approach (Xu et al. 2012). Oberkampf and Helton (2001) had suggested the need for a modelling approach that will take into account epistemic and aleatory uncertainty in WLC separately, so as to avoid underestimating the overall impact of uncertainty. Scope et al. (2016) identified the sources and nature of parametric uncertainty. They grouped multiple sources of data, estimation error, inherent randomness, and data gaps under aleatory; linguistic vagueness, ambiguity, imprecision, lack of data, and unrepresentative data under epistemic; and data collection error under both aleatory and epistemic. Aughenbaugh and Paredis (2006) described how interval or probability bounds analysis can be used to separate aleatory and epistemic uncertainties.

RESEARCH METHOD

Gregg, Kulkarni and Vinz´ (2001) stated that, the development of an information system based on a concept that is rooted in literature, and which solves a problem in a new way at organisational or individual level constitute a research. They noted that, traditionally, research in information systems takes up either a positivist/post positivist or interpretive/constructivist approach. Recent development in information technology makes these paradigms unsuitable for Information systems-based research. This is because the explanations, justifications and methods associated with the paradigms are insufficient for software engineering researches (Deng & Ji, 2018).

Gregg, Kulkarni and Vinz´ (2001) proposed a 3-stage framework for information system research, the philosophical paradigm of which they called 'socio-technologist/developmentalist paradigm'. They stressed that, while positivist/postpositivist paradigm focuses on the confirmation of prepositions, the interpretive/constructivist paradigm is concerned with the generation of new concepts and the socio-technologist/developmentalist paradigm is concerned with

the creation of new ideas. The stages of the proposed framework are conceptualisation, formalisation and development.

The conceptual stage is the first and most important stage. This stage involves defining the research problem by grounding it in literature to generate theoretical constructs and define the needs and specification for the research effort. This stage is very important that none of the two subsequent stages can stand as a research without it.

The formalisation phase involves a systematic description of the system concept using mathematical and/or logical technique which helps to avoid misconceptions and misunderstanding about the system. Tools that could be used include flowcharts, unified modelling language, heuristics etc.

The developmental stage involves the generation of prototypes as a proof of concept, proof by demonstration to test the validity of a proposed solution in a controlled environment. Prototyping is an iterative process with each subsequent stage seeking to improve on the success of a previous stage. It aims to assess the feasibility of the design and test the functionality of the proposed system. Typical activities at this stage includes: mathematical modelling and evaluation, math/logic proofs, analytical and numerical modelling, and computational analysis.

This research adopts the 3-stage information systems methodology framework proposed by Gregg, Kulkarni and Vinz´ (2001). Figure 1 is a summary of the research method adopted. The research begins with problem identification narrated in the introduction section. A literature review was then conducted to articulate types, purpose and classifications of cost models, data requirement, data format and transformation, and risk and uncertainty in WLC, which then provides the basis for identifying framework component, processes, constructs and variables that were then synthesized using basic flowchart symbols to design the conceptual framework. It should be noted that, the development stage of the research process is outside the scope of this paper as the intent is to propose a conceptual framework as a first step.

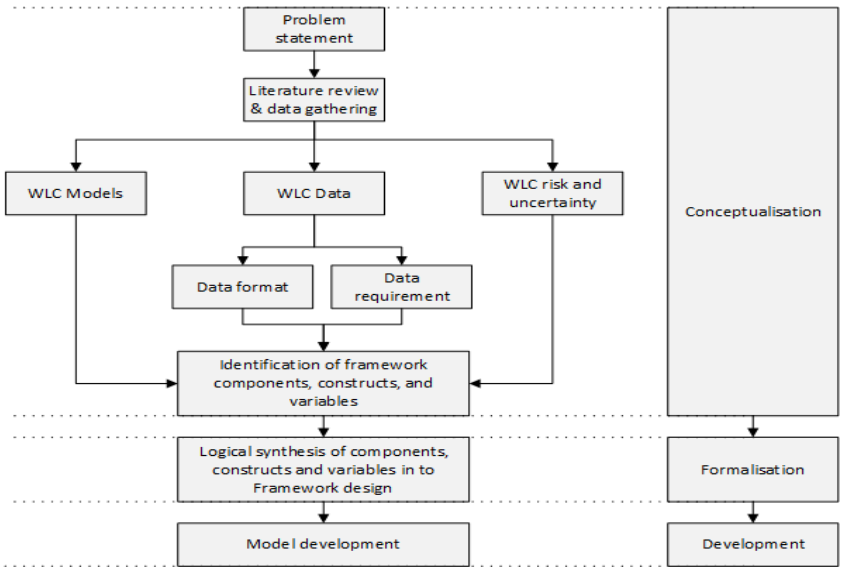


Figure 1:Research Process

RESULTS AND DISCUSSION

Model selection framework

Figure 2 shows a flowchart representation of the conceptual framework.

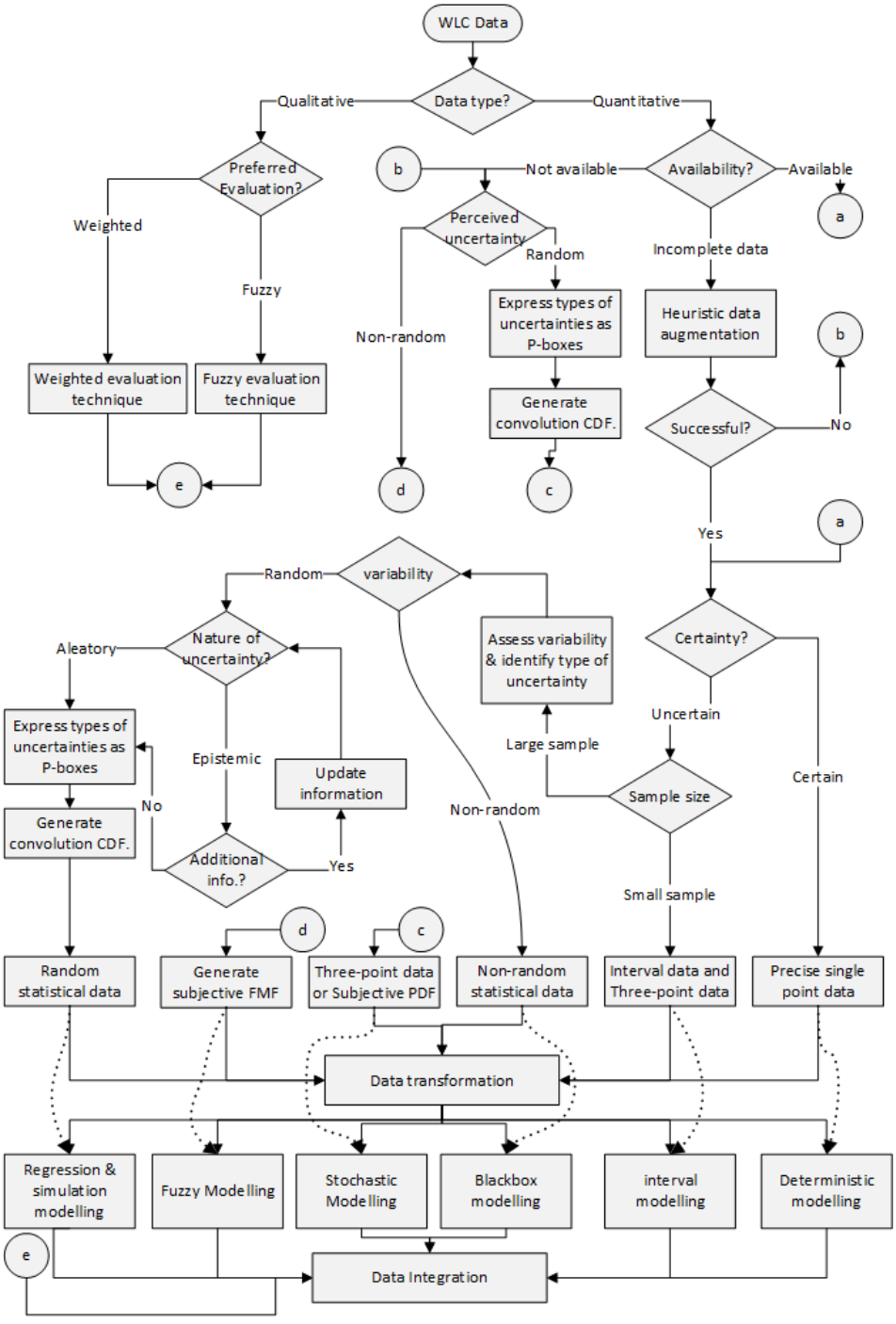


Figure 2: Conceptual framework for data transformation and model selection

In the figure, oval, rectangle, and diamond shapes represents beginning/end of the process, operations, and decisions respectively. Solid arrows show the direction of flow of activities, while dotted arrows show the direction of mapping of data type to modelling technique. The framework portrays seven different scenarios that relates seven quantitative data types to six modelling approaches.

The first scenario is when WLC data is qualitative, the use of standard pair-wise weighted evaluation technique by Dell'Isola and Kirk (2003) or the modified FST-based technique is recommended.

Where the data is quantitative, evaluation is on the basis of availability, certainty, sample size, variability, and nature of uncertainty. The second scenario is where data is available and certain. This scenario is rare in real life given the level of uncertainty associated with WLC data. In this situation however, precise single-point data and deterministic models should be the preferred data format and model respectively.

The third scenario is where quantitative data is available, uncertain, of large sample and random. A frequency distribution should be prepared based on a which a standard PDF can be fitted. If this is done successfully, then, the relevant parametric uncertainties should be identified and assessed as recommended by Goh et al (2010), otherwise the data is treated as non-random.

The fourth scenario is the case where data is available, uncertain, of large sample, and non-random. Non-stochastic analysis involving black-box models like ANN and GA is recommended for this kind of data. Alternatively, fuzzy sets can be generated by classifying and assigning membership values to the data set for use in fuzzy model.

The fifth scenario is where quantitative data is uncertain but of small sample. Small samples of data are not sufficient for constructing a frequency distribution, let alone determining the variability of the data set. This kind of data should be used for interval, three-point estimates, and sensitivity analysis using deterministic models.

The sixth scenario is where quantitative data is available but incomplete. In this situation, data augmentation should be attempted as a first step. Where the data is successfully augmented, then depending on the sample size, scenario four or five can be initiated, otherwise, the situation is treated as the case of unavailable data.

The seventh scenario is where quantitative data is not available. The only option in this situation is to elicit subjective data from professionals. Subjective probabilities or FMF should be obtained from expert opinion and used in stochastic and Fuzzy models respectively.

Data transformation processes

Table 1 is a summary of data types, data levels, data formats, and applicable transformation procedures.

From the table, seven data types, fourteen data formats, and six data levels can be identified. The data types are: Single-Point data (SPD), Range or interval data (RD), Three-Point Data (TPD), Stochastic Three-Point Data (STPD), Fuzzy Data (FD), Probability Density Function (PDF), and Statistical Data (SD).

Table 1: Summary of data transformation

Level	Data format	SPD	RD	TPD	STPD	FMF	PDF
1	SPD	-	-	-		-	-
2	RD	Compute the average	-	-		-	-
3	TPD	Use the likely value	Use the lower & upper values	-		-	-
4	STPD	Use the likely value	Use expanded bound	Use the expanded bounds & likely value			
5	**FD**					-	
5a1	RFMF	Use defuzzified value of FMF	Use the lowest & highest values of the FMF	Use the lowest, highest, & defuzzified value of FMF	Use the expanded bounds, defuzzified value, & confirm the 25% condition		Transform using prob. Pos. consistency principle
5b1	TFMF	As above	As above	As above	As above		As above
5c1	TrpFMF	As above	As above	As above	As above		As above
5d1	GFMF	As above	As above	As above	As above		As above
5	**PDF**						
5a2	OPDF	Use the modal value of the PDF	Use the lowest & highest values of the PDF	Use the range & the modal value	Use the range, modal value, & confirm the 25% condition	Transform PDF into FMF using pro./pos. consistency principle	-
5b2	SPDF						
5b21	UPDF	Use the average value of lower & upper bounds	Use adjusted bounds			Transform PDF into FMF using pro./pos. consistency principle	
5b22	TPDF	Use the likely value	Use adjusted bounds	Use adjusted bounds & the likely value	Use lower and upper bounds, likely value, & confirm the 25% condition	Transform PDF into FMF using pro./pos. consistency principle	
5b23	TrpPDF	Use the mean value	Use range of likely values	Use lower & upper bound and the average of range of likely values	Use lower and upper bounds, average of the likely values, & confirm the 25% condition.	Transform PDF into FMF using pro./pos. consistency principle	
6	**SD**						
6a	RSD	Compute the average.	Use the range.	Use the range and modal value	Use the expanded range, mean value & confirm the 25% condition	Transform PDF into FMF using pro./pos. consistency principle.	Convert data to frequency distribution & fit-in a PDF
6b	NRSD	Compute the average	Use the range	Use the range & modal value	-	Develop a FMF	-

FD has four variants or formats: Rectangular Fuzzy Membership Function (RFMF), Triangular FMF (TFMF), Trapezoidal FMF (TrpFMF), and Gaussian FMF (GFMF). PDF has two main variants, i.e. Objective PDF (OPDF) derived from historical data and Subjective PDF (SPDF). The SPDF has three sub-variants or formats which are: Uniform PDF (UPDF), Triangular PDF (TPDF), and Trapezoidal PDF (TrpPDF). Statistical Data (SD) has Random SD (RSD) and Non-random SD (NRSD) as its variants.

The direction of transformation is chosen such that no additional information is needed during the conversion process. On this basis, six levels of data were observed. The lowest level is the SPD as it cannot be transformed to any other level, followed by RD, TPD, STPD at the 2nd 3rd and 4th levels. Both FD and PDF are at the same 5th level as each data type can be transformed to the other. Statistical Data (SD) is at the 6th and highest level as it can be transformed to five lower levels.

Table 2 is a ranking of the data formats based on the number of transforms they can generate. From the table, RSD is ranked 1st as it generates six number transforms. In contrast, NRSD also a variant of SD ranked 9th as it can generate only four transforms. PDFs and FMFs have the same ranking (2nd) as they can each be transformed to five other formats. They are a special class of data as each data type can be transformed to the other. Unlike the other PDFs, UPDF can only provide three transforms and is ranked 10th alongside STPD.

Table 2: Transformability ranking of data formats

Data format	Abbreviation	Number of transforms	Ranking
Random statistical data	RSD	6	1
Objective probability density function	OPDF	5	2
Trapezoidal probability density function	TrpPDF	5	2
Triangular probability density function	TPDF	5	2
Rectangular fuzzy membership function	RFMF	5	2
Triangular fuzzy membership function	TFMF	5	2
Trapezoidal fuzzy membership function	TrpFMF	5	2
Gaussian fuzzy membership function	GFMF	5	2
Non-random statistical data	NRSD	4	9
Stochastic three-point data	STPD	3	10
Uniform probability density function	UPDF	3	10
Three-point data	TPD	2	12
Interval data	RD	1	13
Single Point data	SPD	0	14

The implication of the ranking is that, data formats with the highest ranking are more flexible, so should be preferred since they provide a wider option for transformation. Thus, SPD should be the least desired while RSD should be the most preferred.

CONCLUSION

A conceptual framework for data characterisation and model selection was proposed. Based on the framework, seven scenarios engendered by different mixes of data attributes, generates a mapping of seven quantitative data types onto six modelling approaches. For each scenario, possible logical decisions in terms of the most befitting data type or format and model or modelling technique to deploy were suggested. Two of the seven scenarios identified are not common in WLC practices. The first involves incomplete data where data augmentation is suggested as a first option prior to describing data as unavailable. The second scenario involves the resolution of data variability into epistemic and aleatory uncertainty which reduces the likelihood of underestimate.

To compliment the model selection process, data transformation procedures were articulated, based upon which transformability ranking for fourteen varieties of data formats was determined. Based on the ranking, the most important data formats are random statistical data, probability density functions and fuzzy membership functions. Generally, the framework provides a guide to built-environment cost professionals for; selecting a suitable data format for WLC in different situations; converting WLC data into different types and subtypes which improves data usability and share ability. It also provides a mapping of the data format to an appropriate model which helps in minimizing model-data mismatch.

Finally, this paper has cleared the ground for the development of transformation algorithms and the design and implementation of a database that would facilitate WLC data storage and retrieval in a variety of formats.

REFERENCES

Al-Hajj, R., Pollock, M., Aoud, G., Sun, M., & Bakis, N., (2001). On the requirement for effective Whole Life Costing in an Integrated Environment. Proceedings of the Construction and Building Research Conference of the RICS Research Foundation (COBRA '2001), 3rd to 6th September, Glasgow Caledonian University.

Anderson, D.R., Sweeney, D.J., & Williams, T.A., (2004). Quantitative methods for business. 9th ed. Ohio, USA: South-western.

Ashworth, A., & Perera, S., (2016). Cost studies of buildings. 6th ed, London: Routledge.

Aughenbaugh, J. M., & Paredis, C. J. M. (2006). The value of using imprecise probabilities in engineering design. Journal of Mechanical Design. 128(4), 969–979.

Boussabaine, A., & Kirkham, R. (2004). Whole Life Cycle Costing: Risk and Risk Responses. 1st ed., Oxford: Blackwell Publishing Ltd.

Bowen, P. A., Walvaardt, J. S., & Taylor, R. G. (1987). Cost modelling: A process modelling approach. In: P.S. Brandon, ed. Building cost modelling and computers. London: E & FN Spon.

Bruns, M. Paredis, C. J. J., & Ferson, S., (2006). Computational methods for decision making based on imprecise information. Proceedings of the NFS Workshop on Reliable Engineering Computing: Modelling Errors and Uncertainty in Engineering Computations, Savannah, Georgia, pp. 1761–1769,

Dell'Isola, A. J., & Kirk, S. J. (2003). Life Cycle Costing for Facilities, 1st Ed, Construction publishers and consultants, US

Dell'Isola, A.J., & Kirk, S. J. (1981). Life Cycle Costing for Design Professionals, 1st ed., New York: McGraw – Hill Inc.

Deng, Q., & Ji, S. (2018). A Review of Design Science Research in Information Systems: Concept, Process, Outcome, and Evaluation. Pacific Asia Journal of the Association for Information Systems. 10(1), 1-36.

El-Haram, M. A., Malcolm, R., Horner, W., & Marenjak, S. (2007). Accuracy of estimating whole life cost: In W., Huges, ed. Proceedings of the inaugural construction management and economics "Past Present and Future 'conference CMER5, 16th to 18th July, University of Reading, UK, pp 527-534.

Ellingham, I., & Fawcett, W. (2006). New Generation Whole-Life Costing: Property and construction decision-making under uncertainty, 1st ed., UK: Taylor and Francis.

Fellows, R., & Liu, A, (2003). Research Methods for Construction. 2nd ed, UK: Blackwell Science Ltd.

Flanagan, R., & Jewell, C. (2005). Whole Life Appraisal for Construction. 1st ed. Oxford: Blackwell Publishing Ltd.

Flanagan, R., Kendell, A., Norman, G., & Robinson, G. D. (1987). Life cycle costing and risk management. Construction Management and Economics, 5(4), 53- 71.

Galway, L. A. (2007). Subjective Probability Distribution Elicitation in Cost Risk Analysis. USA: RAND Corporation.

Gregg, D. G., Kulkarni, U. R., & Vinzé, A. S., (2001). Understanding the Philosophical Underpinnings of Software Engineering Research in Information Systems. Information Systems Frontiers, 3(2), 169–183.

Higham, A. P., Fortune, C. J., & James, H. (2015). Life-cycle costing: Evaluating its use in UK practice. Structural Survey. 33(1), 73-87

Hoar, D. W., (2007). An Overview of life-cycle costing techniques. Journal of Property Management, EmeralBackfiles, 6(2), 92-98.

Huijbregts, M. A. (1998). Application of uncertainty and variability in LCA. International Journal of Life Cycle Assessment. 3, 273–280.

Huijbregts, M. A. J., Gilijamse, W., Ragas, A. M. J., & Reijnders, L. (2003). Evaluating uncertainty in environmental life-cycle assessment. A case study comparing two insulation options for a Dutch one-family dwelling. Environmental Science and Technology. 37(11), 2600–2608.

Ilg, P. Scope, C., Muench, S., & Guenther, E. (2017). Uncertainty in life cycle costing for long-range infrastructure. Part I: levelling the playing field to address uncertainties. International Journal Life Cycle Assessment. 22, 277-292.

Kim, G., An, S., & Kang, K. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Building and Environment. 39, 1235 – 1242.

Kirkham, R. (2015). Ferry and Brandon's cost planning of buildings. 9th ed, Chichester: John Wiley & Sons Ltd.

Kishk, M., & Al-Hajj, A. (2000). A fuzzy model and algorithm to handle subjectivity in life cycle costing bases decision-making. Journal of Financial Management of Construction and Property. 5(1&2), 93-104.

Kishk, M., (2002). Towards effective consideration of non-financial factors in the design and management of construction assets. Journal of Financial Management of Property and Construction. 7(3), 163-173.

Kishk, M., Al-Hajj, A., & Pollock, R. (2002). An innovative integrated approach to whole life costing. Journal of Financial Management of Property and Construction. 7(1), 31-40.

Kishk, M., Al-Hajj, A., Pollock, R., Aoud, G., Bakis, N., & Sun, M. (2003). Whole Life Costing in Construction (A state of the Art Review). London: RICS foundation.

Norman, G. (2007). Life Cycle Costing: Briefing. Journal of Property Management, EmeraldBackfiles. 8(4), 344-356.

Oberkampf, W. L., & Helton, C.J. (2001). Mathematical Representation of Uncertainty, Proceedings of the 19th American Institute of Aeronautics & Astronautics Applied Aerodynamics Conference, 6th -19th April, Seattle, WA.

Opawole, A., Babatunde, S. O., Kajimo-Shakantu, K., & Ateji, O. A. (2020). Analysis of barriers to the application of life cycle costing in building projects in developing countries: A case of Nigeria. Smart and Sustainable Built Environment. 9(4), 503-521.

Ross, T. J. (2004). Fu.zzy Logic with Engineering Applications, 2nd ed. England: John Wiley and Sons Ltd.

Saridaki, M., Psarra,M., & Haugbølle, H. (2019). Implementing life-cycle costing: data integration between design models and cost calculations. Journal of Information Technology in Construction (ITcon). 24, 14-32.

Scope, C., Ilg, P., Muench, S., & Guenther, E. (2016). Uncertainty in life cycle costing for long-range infrastructure. Part II: guidance and suitability of applied methods to address uncertainty, International Journal of Life Cycle Assessment, 21, 1170-1184.

Seeley, I. H. (1998). Building Economics. 4th ed, UK: Macmillan Press Ltd.

Skitmore, M., & Marston, V. (1999). Cost Modelling. London: E & FN SPON.

Spiegel, M.R., Stephens, L. R., (1999). Statistics. 4th ed. USA: McGraw-Hill Inc.

Stephen, L. & Berenson, K. (2005). Statistics for managers, 4th Ed., UK: Pearsons Prenticehall.

Tran, T., Pham, T., Carneiro, G., Palmer, L. & Reid, I., (2017). A Bayesian Data Augmentation Approach for Learning Deep Models. Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), 4th to 7th December, Long Beach, CA, USA.

VanDyk D. A., & Meng, X., (2001). The Art of Data Augmentation. Journal of computational and Graphical Statistics. 10(1), 1-50.

Wang, N., Chang, Y., & El-Sheikh, A. A. (2012). Monte Carlo simulation approach to life cycle cost management, Structure and Infrastructure Engineering. 8(8) 739-746.

Xu, Y., Elgh, F., Erkoyuncu, J.A., Bankole, O. Goh, Y., Cheung, W. M., Baguley, P., et al. (2012). Cost Engineering for manufacturing: Current and future research. International Journal of Computer Integrated Manufacturing. 25(4-5), 300-314.

Zheng, Q., Yang, M., Tian, X., Jiang, N., & Wang, D. (2020). A full stage data augmentation method in deep convolutional neural network for natural image classification. Discrete Dynamics in Nature and Society. 2020, 1-12.

Zimmermann, H. J. (2001). Fuzzy set theory and it's application. 4th ed. Massachusetts, USA : Kluwer Academic Publishers.