Journal of Building Engineering 91 (2024) 109546

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Application of Knowledge Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design

Carmen Llatas[a,*], Bernardette Soust-Verdaguer[a], Luis Castro Torres[b], Daniel Cagigas[b]

[a] *Departamento de Construcciones Arquitectónicas I, Escuela Técnica Superior de Arquitectura, Universidad de Sevilla, Spain*
[b] *Departamento de Arquitectura y Tecnología de Computadores, Escuela Técnica Superior de Ingeniería Informática, Universidad de Sevilla, Spain*

ARTICLE INFO

Keywords:
KDD
Construction database
Triple bottom line
LCSA
Design
BIM
Building

ABSTRACT

Life Cycle Sustainability Assessment (LCSA) can help to predict the impact of products and services, such as buildings, during their entire life cycle. However, it is an extensive data method. The Building Information Modelling (BIM) method can contribute towards reducing the effort involved and simplifying the data collection relating to the building elements. To this end, databases adjusted to the BIM workflow are needed to systematise and harmonise the structure of the environmental, economic and social data of said elements. This paper provides a solution to this problem by presenting an innovative Triple Bottom Line (TBL) database with environmental, economic, and social indicators of building elements to support the triple assessment adapted to the BIM workflow. An analysis employing Knowledge Discovery in Databases (KDD) was performed for the first time on this type of database to better understand the correlations between the dimensions. The key contributions include correlation detection, 83 % of which were direct, which showed that, overall, the environmental (CO_2 emissions), economic (cost), and social (labour) dimensions experience similar growth trends. Strong correlations between economic and social variables were found in 68 % of the cases, followed by those of the economic and environmental (32 %), and social and environmental (18 %) variables. Findings from the correlation analysis between the three dimensions reveal their influence on the type of building system, element and material. Four scenarios were thereby identified in accordance with these correlations, to aid in sustainable decision-making. Various growth trends were detected, which can facilitate the implementation of the LCSA.

1. Introduction

The scientific community highlights the impact that buildings exert on the environment, on the economy, and on society. Climate change constitutes one of the main environmental problems, and the building and construction sector plays a key role, since it is re-

---

Abbreviations: BCCA, Base de Costes de la Construcción de Andalucía; BIM, Building Information Modelling; BP, Building Parameter; CDW, Construction and Demolition Waste; CP, Climate Parameter; EPBD, Energy Performance of Buildings Directive; EPD, Environmental Product Declaration; GWP, Global Warming Potential; IEA, International Energy Agency; ISO, International Organisation for Standardisation; KDD, Knowledge Discovery in Databases; LCA, Life Cycle Assessment; LCC, Life Cycle Costing; LCSA, Life Cycle Sustainability Assessment; RQ, Research Question; SDGs, Sustainability Development Goals; S-LCA, Social-Life Cycle Assessment; TBL, Triple Bottom Line; UN, United Nations.

[*] Corresponding author.

E-mail address: cllatas@us.es (C. Llatas).

https://doi.org/10.1016/j.jobe.2024.109546
Received 22 January 2024; Received in revised form 1 May 2024; Accepted 5 May 2024

sponsible for approximately 40 % of CO_2 emissions worldwide [1]. To overcome this situation and reduce the environmental impact of buildings, the scientific community and policymakers recommend measuring and drastically reducing CO_2 emissions. In this vein, current international sustainability actions, such as the United Nations Sustainability Development Goals (SDGs) [2] and the European Green Deal [3], include ambitious objectives to reduce the effects of climate change. Likewise, the Energy Performance of Buildings Directive (EPBD) [4] proposes the communication of the embodied and operational whole carbon emissions produced by the buildings. To this end, Life Cycle Assessment (LCA) constitutes the most recommended technique, since it offers the most objective and quantitative method for the assessment of the life cycle environmental impacts of any product or service [5].This method has been largely applied to building materials and to completed buildings [6–10].

LCA can also be applied to the economic and social dimensions [11]. It enables the development of Life Cycle Costing (LCC), which assesses economic aspects, and of Social Life Cycle Assessment (S-LCA), which assesses social aspects. The former technique has a longer track record than does S-LCA, which has been applied very little in buildings [12,13]. In search of a more qualitative and objective method to assess the sustainability of buildings, the Life Cycle Sustainability Assessment (LCSA) has been conceived as the sum of the three aforementioned methods (LCA, LCC, S-LCA). This technique has a direct connection with SDGs, for both the assessment of indicators and their monitoring [14]. However, a handicap has been detected in the form of its complexity in data management, especially during the design phase [15–17].

The design phase constitutes a key stage since it is the easiest and cheapest juncture to implement measures to reduce the building's impact [18]. Moreover, the Building Information Modelling (BIM) methodology is useful to simulate these impacts during the design process [15]. This methodology triggered a revolution in the construction sector, once it became mandatory in the European Union in 2018 for public works tenders [19]. Its ability to store and manage data in an automated way during the design stage makes it a powerful methodology to integrate the LCSA [15]. Accordingly, BIM becomes an effective tool in decision-making towards reducing the impact of buildings [15]. In fact, recent studies highlight the benefits of the LCSA and BIM integration where the most relevant challenges are the complexity in considering the triple dimensions in the Life Cycle Inventory phase, the data interoperability, and the software integration [13,15]. However, the integration of LCSA into BIM remains to be fully explored [15].

A major barrier is the lack of specific databases focused on the harmonisation of the three dimensions. The LCSA method needs building information, as well as environmental, economic, and social data regarding the materials, products, and process involved throughout the building life cycle [20]. Hence, it requires the compilation, systematisation, and harmonisation of a large amount of data to develop the life cycle inventory. Even though, in recent years, there has been an increase in the number of multi-dimension building construction databases, they remain largely focused on economic and environmental indicators (e.g., BEDEC [21]). Thus, in the LCSA and BIM implementation, there is a lack of databases that include indicators of all three dimensions and that are adapted to the LCSA workflow in BIM [22]. This means that the information should be structured according to life cycle modules proposed in the ISO 21931-1 standard [23], and to the data granularity of the BIM model.

To address this requirement, recent studies [15,20,22] have proposed a BIM-TBL (Triple Bottom Line) database following the aforementioned principles. It is focused on the LCSA integration into BIM, but remains limited since it considered only a few case studies and a small number of building systems (e.g., structure, envelope). To overcome these limitations, this study presents an enriched version of the BIM-TBL database. It covers the most relevant building elements and includes main groups, elements, and items described in the Andalusian Construction Costs Database, BCCA [24]. This would enable more versatility since it could support any BIM model that includes any of these building elements.

Furthermore, one of the main utilities of any database involves the data analysis and the identification of the trends, together with guidance regarding the decision-making process. To this end, the KDD (Knowledge Discovery in Databases) process provides a useful method to extract "hidden knowledge" [25] from any database. It accesses existing data that cannot be directly transformed into information by the database management software. Patterns, correlations, relationships, and anomalies can all be detected [25]. For this purpose, KDD has been applied in studies related to the building design with several approaches, in order to resolve design clashes in building information models [26] and to discover patterns using KDD and semantic data modelling to establish design patterns [27]. However, its application to databases holding information on building materials remains scarce.

To the best of the authors' knowledge, hitherto there has been no KDD analysis applied to this type of database to detect patterns between the economic, social, and environmental data of the products and materials that make up said database. This constitutes the main research gap detected herein. To fill this gap, this paper applies the KDD method to the BIM-TBL database to analyse the relationships and dependencies between different subtypes or classes of data or information, in an effort to find answers to the following research questions (RQ).

RQ1. Is it possible to find relationships between environmental, economic, and social data of building elements and if so, which aspects mainly affect their correlations?

RQ2. If relationships exist, what are these correlations like, and which dimensions are the most strongly related?

The aim of this study is to detect the correlations between the environmental, economic, and social dimensions of the building items that are used in the LCSA application and based on a BIM workflow.

2. Literature review

2.1. LCA-based construction databases

The LCSA implementation in the building sector involves the collection and systematisation of a large amount of data for the life cycle inventory stage. To this end, the databases that include systematic and predefined values can reduce the effort in the data acquisition for the calculation of the environmental, economic, and social impacts. In the construction field, mainly the economic values for cost estimations are held on construction databases. For example, in Andalusia, the BCCA [24] is a widely used economic database for project budgeting in professional practice, in addition to the existence of BEDEC [21] and Cype [28] in other regions of Spain. These economic databases mainly consider the cost of the acquisition of materials and their installation in the building. However, other life cycle phases, such as the use and maintenance phase or the end-of-life phase (transport process for waste processing using estimated distances) are partially integrated, since they are not fully adapted to the LCSA implementation.

Subsequently, in recent years, mixed databases have appeared, enriched with environmental indicators and impact categories. For example, in Spain, BEDEC [21] includes, in addition to the economic cost of the product acquisition and commissioning phases, the CO_2 emissions during the production and construction phases, and the amount of construction and demolition waste (CDW) generated during the construction phase. However, the approach is focused on the construction process, without including structured information on the building elements during their life cycle. For example, there is a lack of information related to the use stage or to the future demolition waste in the end-of-life stage, which is relevant for the LCSA application.

The LCSA is a data extensive method, which not only requires environmental, economic, and social data but also the harmonisation in the inventories for the assessment of the different dimensions and the building elements, materials, and process data acquisition. For this reason, the use of BIM methodology in the LCSA workflow can contribute towards reducing the effort exerted in data acquisition and its systematisation [15]. Recent LCSA implementations to buildings in BIM underline the difficulties regarding data collection on the environmental, economic, and social dimensions. Several are based on different data sources, combining primary (e.g., surveys) and secondary (e.g., generic data and data obtained from Environmental Product Declarations (EPDs)) data sources, depending on the dimension. For example, Figueiredo et al. [29] used three different data sources for the LCSA implementation: Tally application (Gabi database, a generic environmental database) for the environmental dimension, SINAPI for the economic costs, and primary data based on surveys for the social dimensions. Other studies, such as Salehabadi et al. [30], combined different data sources, such as ecoinvent [31] for environmental data, RSMeans 2019 [32] for economic data, and the Social Hotspots [33] for the social dimension. However, not only do these lack harmonisation of the data structure for the environmental, economic, and social aspects of the building elements, but they also suffer from the exclusion of the modularity principle to organise the building information according to the ISO-21931-1 [23].

In the LCSA, it is important to define a common data structure to organise the information on materials and processes that comprise the inventory [15,22]. On the other hand, other studies [15,20,22]propose a harmonised data structure to consider the triple-dimension assessment adapted to the BIM workflow. In this vein, the limitation of these studies [20,22,29,30,34,35] is that they all involve method and tool validation in case studies and the LCSA data fails to go beyond only a few materials and systems. Nor do these studies focus on analysing the implications of each dimension in the design phase. Moreover, the studies [20,22,29,30,34,35] provide evidence that the information regarding the economic and environmental dimensions is on a higher level of development, while the social dimension is still insufficiently systematised. One of the existing examples is the Social Hotspot database [33], which has limited application to the building sector, and includes limited element flows from limited regions to compare different materials and construction solutions, such as those shown in Salehabadi et al. [30].

To this end, this study focuses on developing a database, integrating information regarding the economic, environmental, and social dimensions, and on providing a harmonised data structure to support the application of LCSA to various building systems and elements. A better understanding of the correlations of the three dimensions in the building design process is also sought.

2.2. Studies on the application of KDD to buildings and construction databases

Data analysis, and in particular Data Mining techniques generally associated with BIM technology, have enjoyed a major upswing in their involvement in architecture and construction in recent years. Data Mining for sustainable construction has been indicated as a trend in the future development of the construction industry [36]. In Zhao et al. [37], for example, the costs of the implementation and management of BIM tools and all the information generated by this type of technology were analysed. Another study [26] employed these techniques to improve productive design processes related to various aspects of building construction. In Petrova et al. [38], a decision support environment based on BIM, Data Mining, and Semantic Data Modelling was proposed. Other studies, such as Pan et al. [39], used BIM as a source of data analysis, where a pattern analysis in BIM models was proposed in order to improve productivity and performance in construction. Similar studies, such as Pan et al. and Pan et al. [40,41], analysed the patterns of files generated by BIM tools in order to discover patterns and design trends that would help in the decision-making of groups of designers. Furthermore, He et al. [42] analysed the correlations between the Climate Parameters (CPs) and Building Parameters (BPs) using Data Mining technology.

Data analysis and Data Mining techniques were also applied to specific environmental aspects of construction such as energy efficiency. In Qian et al. [43], a Data Mining process was carried out with the aim of identifying trends and building design strategies in relation to climatic factors in order to be energy efficient. The KDD process has also been used in the field of construction economics to identify the causes of construction activity delays [44], and cost overruns [45], as well as to investigate social aspects such as to extract useful knowledge from labour resources [46]. However, approaches towards the processes associated with the implementation of building LCA and LCSA are lacking, as are those regarding the correlation between the economic, social, and environmental dimensions, to support the design decision-making using BIM methodology.

3. Methodology

The methodology proposed herein focuses on the data collection and classification, and also on Data Mining. It has been designed to achieve the research objectives through six steps, as shown Fig. 1: 1) Database design and development, 2) Data selection, 3) Data pre-processing, 4) Data transformation, 5) Data Mining, and 6) Interpretation.

Step 1: The methodology started from a BIM-TBL database whose principles, design, and structure were developed in Refs. [22,47], to apply the LCSA of buildings in an automated way in BIM. The database included the data collection regarding the building construction process from the BCCA [24] enriched with the environmental, economic, and social data, as summarised below. It considered the principle of modularity for the building information organised according to the ISO 21931–1 [23]. The database was structured following the classification system and Systematic Building Decomposition [34] of the BCCA [24], including the necessary building materials, machinery, and labour for the construction of different building elements.

The BCCA [24] is a construction database that organises the building information for cost estimation purposes. It uses a hierarchical structure to identify the main systems, (called chapters), the main group of elements (called sub-chapters), the main building elements or working units (called items), and finally to facilitate the decomposition of said items into building materials, machinery, and labour. The main utility of BCCA items involves the composition of the LCSA information on BIM objects. For example, in order to develop the LCSA information of the BIM-object required to evaluate the A1-A3 (product stage) and A5 (construction-installation) information modules of a beam, it will be necessary to use the BCCA items for the beam construction process including the systematic decomposition of the building materials, machinery, and labour. This study is therefore largely focused on the item level.

The information included in the database considered one impact category per dimension. The environmental category included the impact category Global Warming Potential (GWP), measured in units of a kg CO_2 equivalence emissions indicator, since this is the most evaluated environmental impact category in LCA studies [15]. The economic category included the costs in euros, and the social dimension included the employment indicator in working hours since it is the most widely used inventory indicator for S-LCA [48]. Environmental data for the product (A1-A3 modules) and construction (A5 module) stage was estimated by assigning the unit process (activity name) extracted from ecoinvent v3.7.1 [49], the most widely used environmental database for generic data, to the building materials of the BCCA [24]. Economic and social data was estimated using costs and working-hour values from the BCCA [24]. The supplementary data section 1 includes Table S1 with the BCCA items, the breakdown of materials, labour, and machinery and the unit process (from ecoinvent v3.7.1 [49]) used in the calculation.

To this end, the BIM-TBL database was implemented in MS Access [50] as a relational database architecture. It incorporated all the information at different levels of decomposition (from item to sub-element and material level) to perform the automatic assignment of the environmental, economic, and social values to the BCCA items.

Step 2. Once the database was developed, the data was selected and classified. First, a selection of the most relevant construction systems was conducted. It included those that potentially generate the greatest impacts [20], such as foundation, structure, masonry, installations, finishings, envelope, and partitions. Those that generate the lowest impacts [20], such as equipment, waste management process, urban equipment, security, and health, were not included in this study.

The second selection criterion was the functional unit classification to provide a fair comparison of the data. The information regarding the BCCA [24] items included different types of units for the quantity estimations. For example, the beam construction is measured in “m^3”, while the wall construction is measured in “m^2”. This process implied the grouping of the items that have the same functional unit, while maintaining their classification in chapters and sub-chapters of information. Therefore, the selected BCCA items were classified into different functional units using the unit of measure (e.g., kg, m, m^2, m^3), to provide a fair comparison of building elements with similar functions in the system to which they belong.

This stage also focused on analysing the data comparability of different functional units. For example, the unit of measure “unit” was deleted from the analysis due to the great differences in the product characteristics and functions, which could lead to errors or misinterpretations. For example, a “unit” could be a large element such as a lift or a small and insignificant element such as a screw. The selected chapters therefore included only those items that were measured using kg, m, m^2, and m^3.

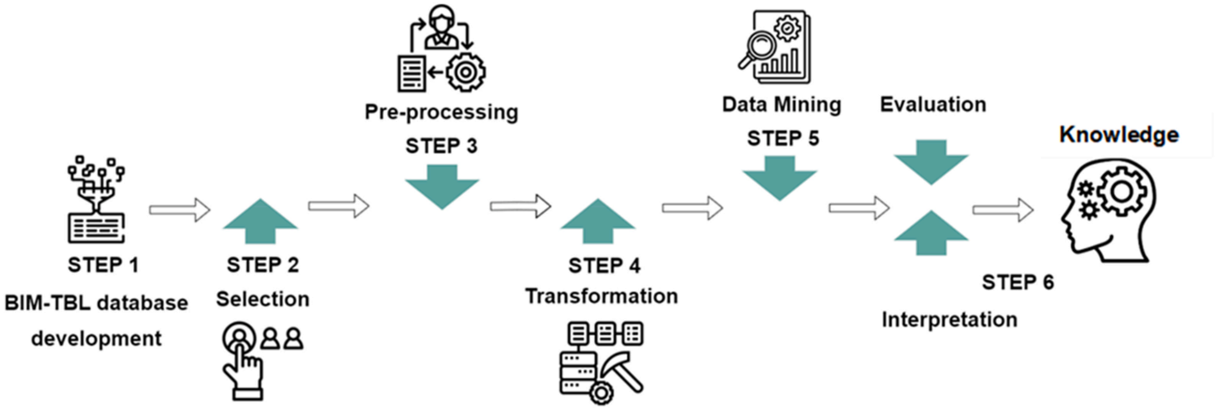


Fig. 1. Flowchart showing the methodology followed in this study.

Step 3. In the pre-processing stage, the data was prepared and transformed for further analysis. The social, economic, and environmental data for each BCCA item was obtained by summing the data from its modules included in the BIM-TBL database (A1-A3, C1, C2, C4). Thus, certain BCCA items were deleted: those with inconsistent data, with values out of range, or with zero values for environmental, economic, or social dimensions. All the values were updated, corrected, and normalised to enable the comparison of the different dimensions (environmental, economic, and social) in later steps. The normalisation of the data implied the adjustment of values measured using different units (e.g., kg CO_2 eq., euros, working hours) to a common scale to enable their comparison. The results of the statistical analysis were presented in charts and carried out in Excel.

Step 4. At this stage, the data was transformed into an appropriate format to proceed with its analysis. In this case, this transformation consisted of the generation of data files in the form of a matrix and implemented in the form of.csv files. These files came from queries made on the MS Access database.

Step 5. This stage allowed obtaining useful information through a series of statistical analysis techniques and algorithms. The goal of this analysis was to seek possible interrelationships between the economic, environmental, and social data of the integrated database. To this end, a bivariate analysis was applied in the form of a correlation between economic, environmental, and social factors. The correlation coefficients between dimensions were denoted as follows.

"r_1" is the correlation coefficient between economic and social dimensions.
"r_2" is the correlation coefficient between economic and environmental dimensions.
"r_3" is the correlation coefficient between social and environmental dimensions.

Three types of correlations were used: Pearson [51], Kendall [52], and Spearman [53]. Pearson's coefficient measures the linear relationship between two continuous variables, and it is appropriate when the variables do not have a normal distribution or when the data has outliers. Kendall's and Spearman's correlation coefficients measure the strength and direction of association that exists between two variables measured on at least an ordinal scale.

The three correlation coefficients were classified as strong, moderate, or weak based on the absolute value of the coefficient. A coefficient between 0.50 and 1.00 indicates a strong direct correlation, while a coefficient between −0.50 and −1.00 indicates a strong inverse correlation.

Step 6. Lastly, in this stage, the interpretation and evaluation of the relationships between the various factors, correlation coefficients, and data visualisation were implemented.

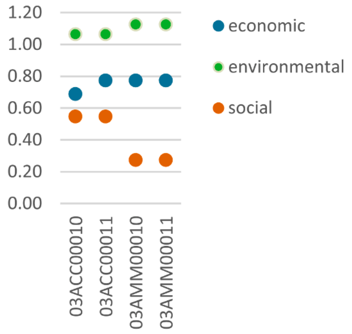
4. Results

4.1. Selected BCCA items

According to the selection criteria and the pre-processing stage, 1566 BCCA [24] items were selected from the BIM-TBL database, and were classified following eleven main BCCA [24]chapters (building systems): earthwork (#2), foundations (#3), structure (#5), masonry (#6), roofing (#7), insulation (#9), finishings (#10), carpentry (#11), glass (#12), painting (#13), and exterior fittings (#15). They were also classified into four main functional units: "kg", "m", "m^2", and "m^3". The graphs in the results and supplementary data section 2 show these items (Figures S1.1 to S1.22).

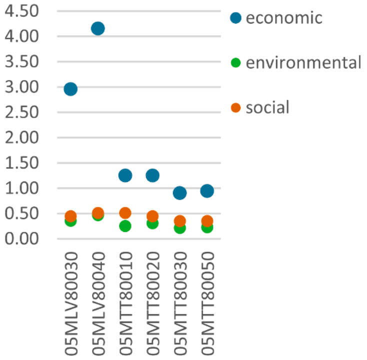
4.2. Database statistical analysis

Statistical analysis shows the data extracted from the database queries. In this vein, to obtain consistent and comparable results, the data was pooled according to the BCCA chapter (building system) (e.g., foundations, structure, roofing), the type of functional unit (including kg, m, m^2, and m^3), and the BCCA item type (e.g., 1 m^3 of pile, 1 m^3 of beam, 1 m^2 of window). Fig. 2a–d shows a selection of the results obtained and include each of the functional units comprised in this study (kg, m, m^2, and m^3). The full set of charts is provided in the supplementary data section 2 (Figures S1.1 to S1.22).



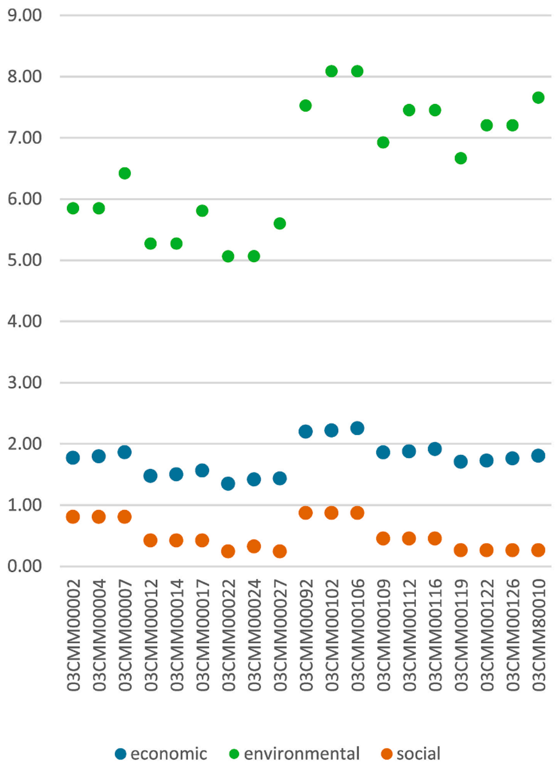
BCCA item code	Description
03ACC00010	Steel in rebar B 400 S in foundation elements
03ACC00011	Steel in rebar B 500 S in foundation elements
03AMM00010	Steel in electro-welded mesh made with ME B 500 S corrugated wires
03AMM00011	Steel in electro-welded mesh made with ME B 500 T drawn wires

Fig. 2a. Normalised values for Chapter 3 Foundations (Functional Unit "kg").



BCCA item code	Description
05MLV80030	Glued laminated conifer timber GL-24 with adhesive type 1, in straight beams, pieces up to 5 m
05MLV80040	Glued laminated conifer timber GL-24 with adhesive type 1, in straight beams, pieces between 5 and 10 m
05MTT80010	Vertical support of treated pine wood 20x20 cm, for a maximum height of 4 m
05MTT80020	20x20 cm treated pine wood beam
05MTT80030	20x20 cm treated pine wood purlin
05MTT80050	17x20 cm treated pine wood joists

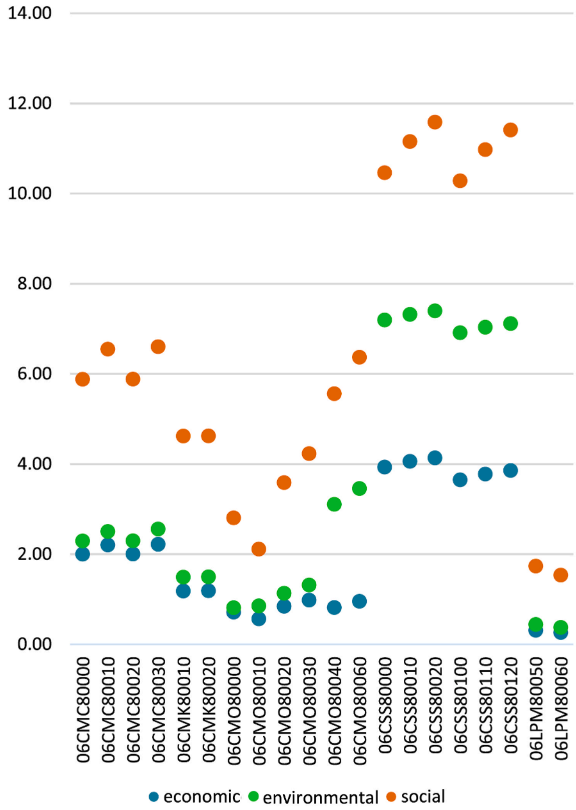
Fig. 2b. Normalised values for Chapter 5 Structure (Functional Unit “m”).



BCCA item code	Description
03CMM00002	45 cm-thick CCP diaphragm wall with mechanical grab 6 m in depth (HA-25)
03CMM00004	45 cm-thick CCP diaphragm wall with mechanical grab 6 m in depth (HA-30)
03CMM00007	50 cm-thick CCP diaphragm wall with hydraulic grab 6 m in depth (ha-30)
03CMM00012	45 cm-thick CCP diaphragm wall with mechanical grab, 6 to 15 m in depth (HA-25)
03CMM00014	45 cm-thick CCP diaphragm wall with mechanical grab, 6 to 15 m in depth (HA-30)
03CMM00017	50 cm-thick CCP diaphragm wall with hydraulic grab, 6 to 15 m in depth (ha-30)
03CMM00022	45 cm-thick CCP diaphragm wall with mechanical grab and more than 15 m in depth (HA-25)
03CMM00024	45 cm-thick CCP diaphragm wall with mechanical grab and more than 15 m in depth (HA-30)
03CMM00027	50 cm-thick CCP diaphragm wall with hydraulic grab and more than 15 m in depth (HA-30)
03CMM00092	60 cm-thick CCP diaphragm wall with hydraulic grab, 6 m in depth (HA-30)
03CMM00102	65 cm-thick CCP diaphragm wall with mechanical grab (HA-25)
03CMM00106	65 cm-thick CCP diaphragm wall with mechanical grab (HA-30)
03CMM00109	60 cm-thick CCP diaphragm wall with hydraulic grab, 6 to 15 m in depth (HA-30)
03CMM00112	65 cm-thick CCP diaphragm wall with mechanical grab, 6 to 15 m in depth (HA-25)
03CMM00116	65 cm-thick CCP diaphragm wall with mechanical grab, 6 to 15 m in depth (HA-30)
03CMM00119	60 cm-thick CCP diaphragm wall with hydraulic grab, more than 15 m in depth (HA-30)
03CMM00122	65 cm-thick CCP diaphragm wall with mechanical grab, more than 15 m in depth (HA-25)
03CMM00126	65 cm-thick CCP diaphragm wall with mechanical grab, more than 15 m in depth (HA-30)
03CMM80010	70 cm-thick CCP diaphragm wall with mechanical grab, more than 15 m in depth (HA-30)

Fig. 2c. Normalised values for Chapter 3 Foundations (Functional Unit “m^2”).

Fig. 2a–d shows that increases do not occur in a linear fashion in each of the dimensions, chapters, and items of the BCCA. On the one hand, several likely patterns of behaviour were detected in similar products and materials. For the functional unit “kg” in Chapter 3 Foundations, Fig. 2a shows that similar tendencies are detected in the same type of product such as steel bars and electro-welded mesh. However, for the functional unit “m” (linear metre) in Chapter 5 Structure, Fig. 2b shows that the dispersion of economic values is greater for timber beams (Glued Laminated Timber-GLT) than for pine wood beams, where the proportion of growth of economic values is much higher than that of pine wood.



BCCA item code	Description
06CMC80000	One-sided limestone masonry
06CMC80010	Two-sided exposed limestone masonry
06CMC80020	One-sided granite stone masonry
06CMC80030	Two-sided granite stone masonry
06CMK80010	One-sided limestone masonry
06CMK80020	One-sided granite stone masonry
06CMO80000	Ordinary one-sided limestone masonry, with perpendiculars for transversal bracing and shingles
06CMO80010	Ordinary limestone masonry for cladding, with perpendiculars for transversal bracing and wedge stones
06CMO80020	Ordinary one-sided limestone masonry, with perpendiculars for transversal bracing and coining rubble
06CMO80030	Ordinary two-sided exposed limestone masonry
06CMO80040	Ordinary one-sided granite stone masonry, with perpendiculars for transversal bracing and coining rubble
06CMO80060	Ordinary two-sided granite stone masonry
06CSS80000	One-sided sawn limestone ashlar masonry
06CSS80010	One-sided bush-hammered limestone ashlar masonry
06CSS80020	One-sided limestone ashlar masonry
06CSS80100	Sawn granite ashlar masonry
06CSS80110	Bush-hammered one-sided granite ashlar masonry
06CSS80120	Honed one-sided granite ashlar masonry
06LPM80050	Two-foot-thick masonry of rough perforated brick, 24x11.5x7 cm
06LPM80060	Two-foot-thick masonry of rough perforated brick, 24x11.5x10 cm

Fig. 2d. Normalised values for Chapter 6 Masonry (Functional Unit “m^{3}”).

On the other hand, in Fig. 2c for the functional unit “m^{2} “in Chapter 3 Foundations, the patterns of behaviours were not as similar for the same type of element (diaphragm wall), where the environmental values show the greatest dispersion. Fig. 2d for the functional unit “m^{3}” in Chapter 6 Masonry shows an even more cluttered pattern of behaviour. However, similar items, such as “Two-foot-thick masonry” (06LPM80050 and 06LPM80060 items), and “One-sided limestone or granite masonry” (06CMK80010 and 06CMK80020) had similar dispersions.

The rest of the analysed values, which included the remaining items grouped according to the other functional units in the chapters of earthwork (#2), foundations (#3), structure (#5), masonry (#6), roofing (#7), insulation (#9), finishings (#10), carpentry (#11), glass (#12), painting (#13), and exterior fittings (#15) are provided in the supplementary data section 2 (Figures S1.1 to S1.22). Similar tendencies were detected therein, where the pattern of behaviour generally depends on the type of building element (function) and the main material. For example, for the functional unit “m^{3}” in Chapter 5 Structure, a different trend was observed for concrete elements (such as pillars, slabs, and beams) to that for timber elements (e.g., pillars) (see Figure S1.21 in supplementary data).

4.3. Knowledge Discovery in databases analysis

Table 1 shows the correlation coefficients for the functional units analysed and the relationships found between the different factors and data in the database, in which correlation indices greater than +0.50 or less than −0.50 are marked in bold. The items included in each functional unit considered those whose functional units are used in each chapter (building system) of the BCCA for the measurement of each type of item: i) for the functional unit “kg”, items from the chapters on foundations (#3) (e.g., steel meshes) and structure (#5) (e.g., steel profiles); ii) for the functional unit “m”, items from the chapters on foundations (#3) (e.g., piles), structure (#5) (e.g., wooden beams), masonry (#6) (e.g., lintels), and roofing (#7) (e.g., roof overhangs); iii) for the functional unit “m^{2}”, items from the chapters on foundations (#3) (e.g., retaining walls), structure (#5) (e.g., slabs), masonry (#6) (e.g., brick walls), roofing (#7) (e.g., tiled roofs), insulation (#9) (e.g., façade insulation), finishings (#10) (e.g., flooring), carpentry (#11) (e.g., windows), glass (#12) (e.g., polycarbonate panelling), painting (#13) (e.g., façade painting), and exterior fittings (#15) (e.g., ground cover vegetation); and iv) for the functional unit “m^{3}” items from the chapters on earthwork (#2) (e.g., transport of soil), foundations (#3) (e.g., footings), structure (#5) (e.g., concrete beams), and masonry (#6) (e.g., stone walls).

As shown in Table 1, the relationships found at the functional unit level have a strong positive correlation between the economic factor “total cost” and the social factor “labour hours” (r_1), in all the functional units except for units “m” and “m^{2}”, since in those

Table 1
Correlation coefficients obtained per functional unit.

Functional unit		Kendall			Pearson			Spearman		
		Econ.	Enviro.	Social	Econ.	Enviro.	Social	Econ.	Enviro.	Social
kg	Econ.	1.00			1.00			1.00		
	Enviro.	0.10	1.00		0.37	1.00		0.14	1.00	
	Social	**0.82**	−0.02	1.00	**0.91**	−0.00	1.00	**0.93**	−0.08	1.00
m	Econ.	1.00			1.00			1.00		
	Enviro.	**0.68**	1.00		**0.68**	1.00		**0.68**	1.00	
	Social	0.21	0.19	1.00	0.21	0.19	1.00	0.21	0.19	1.00
m^2	Econ.	1.00			1.00			1.00		
	Enviro.	0.32	1.00		0.29	1.00		0.44	1.00	
	Social	0.11	0.07	1.00	−0.08	−0.04	1.00	0.16	0.10	1.00
m^3	Econ.	1.00			1.00			1.00		
	Enviro.	**0.64**	1.00		0.44	1.00		**0.71**	1.00	
	Social	**0.53**	0.28	1.00	**0.64**	0.08	1.00	**0.71**	0.33	1.00

groups the highest number of BCCA items and greatest diversity of BCCA chapters were analysed. In the case of the functional unit “m”, a positive relationship between the social factor “labour hours” and the environmental factor “CO_2 emissions” (r_3) was detected, and even for the unit “m^3” (Kendall and Spearman coefficients). For the functional unit “m^2”, however, no correlation was found between the three factors, since in this group a wide variety of items from eleven chapters were analysed. Table 1 also shows that all the relationships found a match in at least two of the three correlation indices used (Pearson, Kendall, and Spearman).

Given the great variety and diversity of values (Fig. 2a–d), it was found that the identification of correlations may have limitations if the values present different patterns of behaviour. Therefore, the level of disaggregation of the information was increased to detect possible correlations within the chapters of the BCCA [24].

Fig. 3a–d shows the correlation matrix of the results obtained including each of the functional units comprised in this study (kg, m, m^2, and m^3). The foundation items are shown to have different correlation coefficients than the structure items. The complete set of charts, with the full list of BCCA chapters and items analysed, is included in the supplementary data section 2 (Figures S2.1 to S2.4).

The results for the kg functional unit show different tendencies depending on the chapter. For instance, (see Fig. 3a), the correlation between economic and social factors (r_1) was strong and positive (0.85–0.95) for the structure but weak and negative (−0.58) for the foundations, while the social and environmental correlation (r_3) was negative (−1.00) for the foundations and positive, although hardly related (0.06–0.12), for the structure. Economic and environmental correlations (r_2) are positive in both chapters, although they remain stronger in the foundations chapter. Furthermore, the three correlation indices (Kendall, Pearson, Spearman) provide the same values for all the dimensions related to the items in the foundations chapter of the BCCA, and the social and environmental correlations have a perfect negative correlation with a value of −1.00.

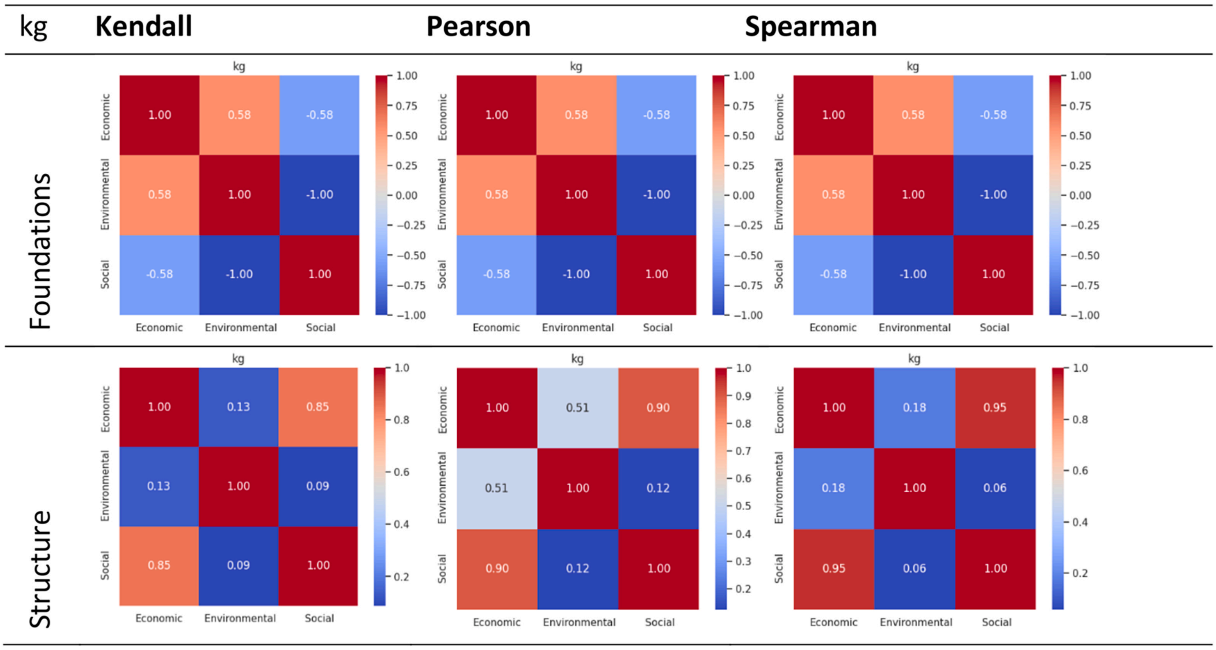


Fig. 3a. Correlation matrix for the selected BCCA items measured in “kg”.

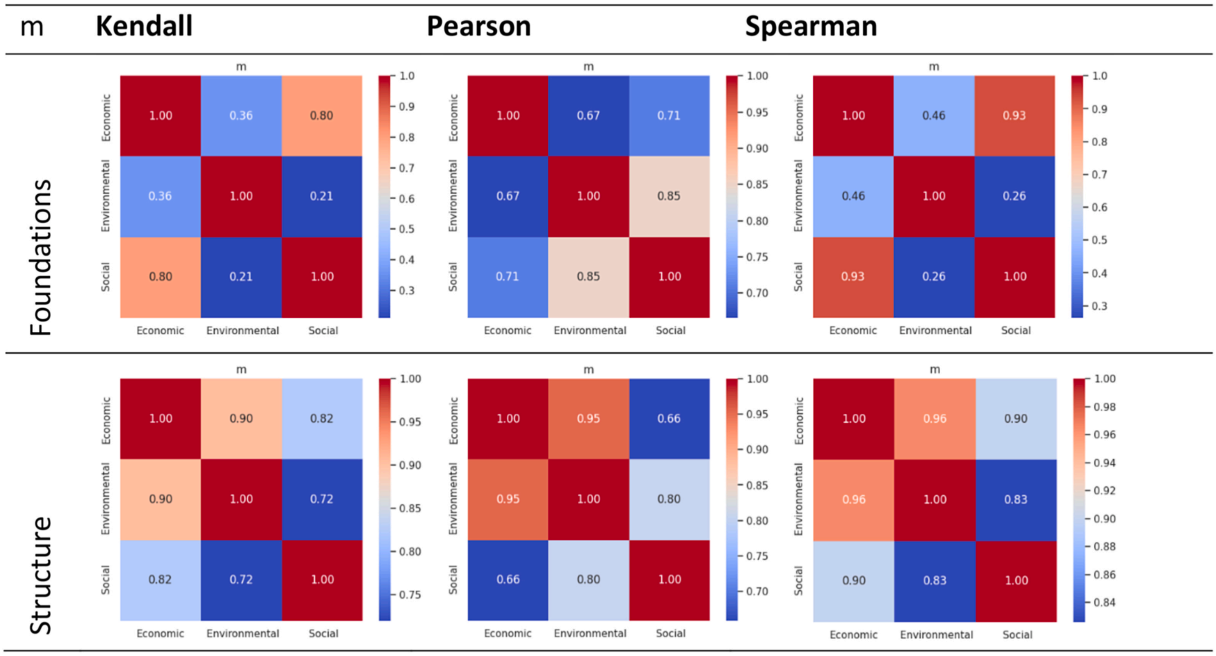


Fig. 3b. Correlation matrix for the selected BCCA items measured in "m".

The results for the functional unit "m" (see Fig. 3b) show similar strong positive correlations between the economic and social dimensions (r_1) for the foundations (0.71–0.93) and structure (0.66–0.90) chapters. However, the correlation between the economic and environmental dimensions (r_2) as well as the social and environmental dimensions (r_3) are stronger (0.90–0.96) in the structure chapter than in the foundations chapter (0.36–0.67).

Fig. 3c shows the results for BCCA items in the foundations and structure for the functional unit "m^2". Depending on the building element and material, the correlation matrix shows different tendencies. In the concrete slabs, all the correlation indices are positive and economic and social dimensions are strongly related, while the concrete and steel slabs show a negative weak correlation between the economic and social dimensions (−0.11, −0.61). In the foundation items, the most strongly related dimensions are economic and environmental (0.76–0.89), followed by economic and social (0.61–0.78), while in the concrete slabs, the most strongly related dimensions are economic and social (0.74–0.99) followed by economic and environmental (0.46–0.69). The concrete and steel slabs present strong negative correlation (−0.61) between the economic and social dimensions in the Pearson coefficient, and strong positive correlation (0.57) between the economic and environmental dimensions in the Spearman coefficient.

Fig. 3d shows the results for BCCA items in the foundations and structure chapters for the functional unit "m^3". In the case of the structure chapter, the strongest correlation is found between the economic and social dimensions (r_1), with similar behaviour to that of the functional units "kg" and "m^2" for concrete slabs. In the case of the foundations chapter, similar behaviour has been found with respect to the functional unit "m^2" since the strongest correlations are between the economic and environmental dimensions (r_2), followed by the economic and social dimensions (r_1), and subsequently by the social and environmental dimensions (r_3).

The other correlation matrices analysed for the rest of the functional units in the other chapters are provided in the supplementary data section 2 (Figures S2.1 to S2.4). Table 2 summarises the ranges of the correlation coefficients (r_1, r_2, r_3) obtained for the combinations of the three related dimensions, including the number of related items in each functional unit group. The detected tendencies indicate that the majority of the items present a strong correlation between economic and social variables (r_1) (e.g., structure). However, other items, such as those of the finishing chapter, have a weak correlation between economic and social variables (r_1). A lower number of items have a strong positive economic and environmental correlation (r_2) (e.g., foundations). Finally, the strong correlation between social and the environmental dimensions (r_3) presents the lowest incidence.

5. Discussion

5.1. Main contributions and lessons learnt

5.1.1. In answer to RQ1

The contributions of this work lie in the demonstration that it has been possible to identify both positive and negative correlations between environmental, economic, and social variables of the BCCA items analysed, in answer to RQ1. As shown Table 2, most of the correlation coefficients are positive (83 %), as opposed to negative (17 %), which indicates a direct rather than an inverse relationship between the variables analysed. In 39 % of the cases there is a strong correlation (greater than +0.50, or less than −0.50) in all the correlation factors analysed (Kendall, Pearson, and Spearman), whereby 8 % is negative and 92 % positive, with non-existent correlation or weak/moderate correlation in at least one factor in the remaining cases (61 %) (see Table 2).

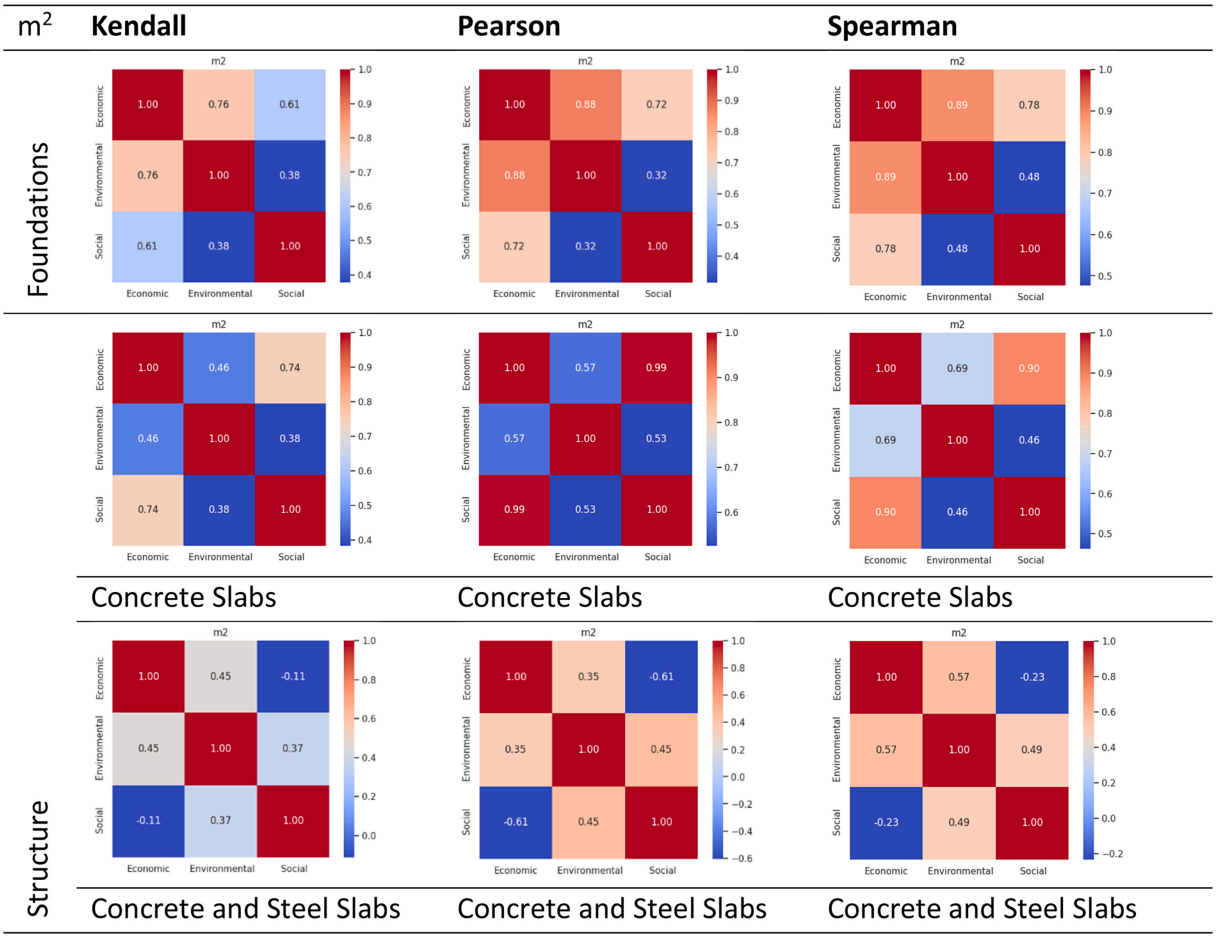


Fig. 3c. Correlation matrix for the selected BCCA items measured in "m^2"

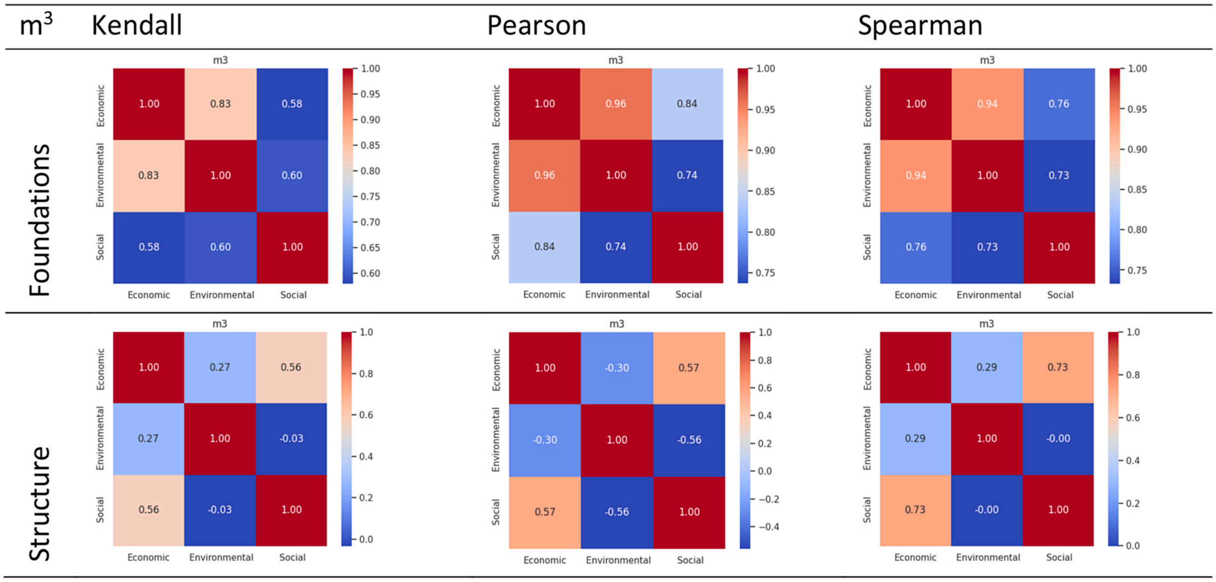


Fig. 3d. Correlation matrix for the selected BCCA items measured in "m^3"

Table 2
Range of correlation coefficients (r) obtained per chapter and functional unit (f). Resulting scenarios[a] (S).

hapter	Functional unit f	Number of items N	Economic and Social r_1	Economic and Environmental r_2	Social and Environmental r_3	Scenario[a] S
2. Earthwork	m^3	17	**0.85–0.93**	**0.81–0.92**	**0.69–0.86**	S1
3. Foundations	kg	4	**−0.58**	**0.58**	**−1.00**	S4
	m	31	**0.71–0.93**	0.36–0.67	0.21–0.85	S4
	m^2	19	**0.61–0.78**	**0.76–0.89**	0.32–0.48	S4
	m^3	81	**0.58–0.84**	**0.83–0.96**	**0.60–0.74**	S1
5. Structure	kg	18	**0.85–0.95**	0.13–0.51	0.06–0.12	S4
	m	7	**0.66–0.90**	**0.90–0.96**	**0.72–0.83**	S1
	m^2	79	**0.74–0.99**	0.46–0.69	0.38–0.53	S4
	m^3	48	**0.56–0.73**	−0.30-0.29	−0.00-(-0.56)	S4
6. Masonry	m	71	0.32–0.57	0.23–0.39	−0.08-0.23	S3
	m^2	155	**0.57–0.71**	0.41–0.57	0.17–0.22	S4
	m^3	20	**0.78–0.90**	**0.63–0.82**	0.49–0.67	S4
7. Roofing	m	67	0.24–0.48	0.21–0.68	0.30–0.52	S3
	m^2	33	**0.61–0.86**	0.26–0.55	0.35–0.66	S4
9. Insulation	m^2	108	**0.56–0.80**	0.34–0.47	0.21–0.40	S4
10.1. Wall finish.	m^2	132	−0.06-0.09	**0.54–0.73**	0.05–0.08	S4
10.2. Floor finish.	m^2	127	−0.04-0.16	0.03–0.13	0.29–0.50	S3
10.3 Ceilings	m^2	26	−0.04-0.30	0.03–0.44	−0.09-0.29	S3
11. Carpentry	m^2	348	0.13–0.44	−0.02-0.04	−0.39-(-0.14)	S3
12. Glass	m^2	66	**0.57–0.82**	0.11–0.55	0.42–0.56	S4
13. Painting	m^2	85	**0.94–0.97**	0.11–0.14	0.14–0.17	S4
15. Exterior fitting	m^2	24	0.39–0.47	0.28–0.51	0.12–0.29	S3

[a] See Section 5.3.1. Contribution to sustainable design.

Correlations were found on similar elements, and sub-elements with similar materials such as steel bars or electro-welded mesh (see Fig. 3a). Negative correlations were detected between social and environmental impacts (r_3), which means that when environmental values increase, social values are reduced. However, the correlations between the various dimensions and the growth of the values were different. Generally, the maximum and minimum of each series have different growth logic, and hence finding relationships between variables was more difficult.

This study shows that the linear correlation (1) is only detected when the same variable is considered, and that the more disaggregated the data analysed is, the more correlations can be found. It was also found that the correlations depend mainly on the functional unit, the type of item (building element) analysed within each building system, and its materials. For example, in the case of foundations and structure, at a general level per “kg” of functional unit, only strong direct correlation (0.82–0.93) was found between economic and social variables (r_1) (see Table 1), and the remaining variables were not strongly related. This means that if social impact values increase, then there is a proportional growth in the economic costs. However, if the same values are analysed disaggregated in terms of chapters (considering foundations and structure separately), then it was possible to find another level of correlations, as shown in Fig. 2a and Table 2. In this case, for foundation items, the economic and social variables (r_1) changed trend, and became inversely related (−0.58), and other variables appeared strongly related, such as the social and environmental dimensions (r_3), with a perfect negative correlation (−1), and the economic and environmental (0.58) dimensions (r_2). Furthermore, in the case of concrete slabs in the structure chapter, different correlation factors were found for unreinforced slabs compared to the same concrete slab reinforced with steel, as shown in Fig. 3c.

5.1.2. *In answer to RQ2*

The study also detected the most closely related dimensions in answer to RQ2. Firstly, these were the economic and social (r_1) dimensions in which 68 % of all the factors (Pearson, Kendall, and Spearman) obtained strong correlations, followed by the economic and environmental (r_2) dimension (32 % in this case), and by the social and environmental (r_3) dimension (18 %), as shown in Table 2. Although correlations between dimensions do not have to follow a cause-effect pattern, these coefficients can differ in terms of.

- (r_1) The economic (cost) and social (labour hours) variables may be directly interrelated since labour cost is a factor with a major impact on the total cost, which comprises of 30–50 % of the overall project's cost [54]. This relationship is therefore strongly positive in the items of almost all the chapters analysed, with certain exceptions (see Table 2). In the case of the foundations (kg), the relationship is inverse since the cost of the material (steel) has a greater weight in the final cost than that of the labour.
- (r_2) The environmental dimension (CO_2 emissions) may have a direct impact on the economic dimension (cost) since higher CO_2 emissions due to energy consumption in manufacturing leads to higher costs. Moreover, the use of auxiliary machinery in these items can also directly affect both variables, economic cost, and CO_2 emissions. This correlation is significant in the items of earthwork (m^3), foundations (in all functional units except m), structure (m), masonry (m^3), and wall finishings (m^2) chapters. These results are in line with those of other studies in which optimised solutions were found that reduced both factors (e.g., Refs. [55,56]).

- (r_3) The social and environmental variables are the least related in this study, as shown in Table 2. Their correlation is significant in the items of the earthwork (m^3), foundations (kg, m^3), where it is inverse for kg, and of structure (m) chapters. In these items, the use of machinery involves labour (social) for its operation, and this in turn has an impact on CO_2 emissions (environmental). In the case of the foundations chapter, the correlation of the items in “kg” is inverse, with a perfect correlation (−1), as the related items have double labour, but half material, and hence the emissions from material manufacture decrease in the same proportion as labour increases. On the other hand, if other variables had been assessed, such as the transport of workers to and from the site, then the relationship between the social and environmental variables would have been much stronger for all labour-related items, due to CO_2 emissions from such transport.

Table 2 also shows how of the 22 groups of items analysed, the majority (45 %) have only one strong correlation (mainly between social and economic dimensions) compared to 9 % with two strong correlations, and 18 % with 3 strong correlations between all dimensions. Nevertheless, special care was taken to ensure that the groupings of data respected the type of element or function and could contribute towards finding correlations more easily.

5.1.3. Seeking correlations in databases

If the various impact dimensions of the BCCA items were compared and the variability were proportional, then more correlation could be found. Hence, in order to find more correlations, special attention must be paid to the selection of values and the way they are organised.

For positive and/or negative correlations to occur, the increase in values should be proportional. One of the potential strategies for the control of the variability of the values involves differences in the quantities of materials. For example, in the event that *“BCCA item X”* includes half the quantity of material as “*BCCA item Y″*, then the derived impacts for the environmental, economic, and social dimensions are incremental and a clear positive correlation is detected. However, that which differentiates the value of some BBCA items from others is not only the variation in the quantity of material, but also other factors with major influence, such as auxiliary materials and energy consumption. In this vein, this study demonstrates that the correlation analysis should also include the analysis of other parameters, such as the material quantities, auxiliary materials, and energy consumption, to better understand its influence in the derivate impacts. Therefore, the use of the same type of material (e.g., HA-30 concrete) and different building elements (e.g., pillars, beams), or different types of materials while maintaining constant the amount of material used (e.g., 1 m^2 of window) can be useful in the analysis of how those variations influence the correlation between environmental, economic, and social impacts.

5.2. Limitations of the study

The use of LCSA in the building sector remains scarce, as do the data sources for its implementation at the building element and building levels [13,15]. The large amount of data required along with the limited systematic data sources to support the LCSA implementation to buildings constitute the main obstacles.

Consequently, the Data Mining method can facilitate decision-making and reduce the complexity of building element data. The results of the correlation analysis indicate that the incorporation of the items that produce the lowest impact in all dimensions (environmental, economic, and social) into the design process is a complicated process. For example, in the data analysed there is no disproportionate or markedly upward growth in all three dimensions at the same time (see Table 2), which can limit the search for correlations between alternative items. Furthermore, the data analysis focused on a limited number of BCCA elements and was organised in terms of functional unit, while leaving other functional requirements of the design out of the analysis (e.g., thermal comfort). Therefore, although the Data Mining method can help towards understanding the behaviour of large amounts of data, it remains crucial that the data is clearly organised (e.g., using similar alternative building elements, using the same materials) and that the method covers small data ranges.

The data analysed in this paper mainly covers the construction and installation phases of the building items. Nevertheless, it is treated as baseline information to be used for the different life cycle stages. For example, the replacement of a window mainly includes the manufacturing and installation of a new window. Thus, similar data can be used both for the construction phase and the use phase, which can be enriched with information regarding the transport, treatment, and final disposal of the building materials. These data could identify building elements that, despite having a high initial cost, have a lower total cost by requiring less maintenance or even by being recovered at the end of their lifespan. Moreover, this study focuses on analysing the embodied impacts and aspects related to the elements, products, and materials that are installed in buildings throughout their life cycle. Given the decarbonisation of the energy mix in the coming years and its relevance in reducing operational impacts related to energy consumption [57], embodied impacts and aspects related to the materials will require special attention in the building design. However, impacts such as energy use during operation are also relevant when conducting a full LCSA. Therefore, the trade-off between embodied and operational impact should be considered in the assessment of the sustainability of the building.

Another limitation detected herein involves the number and scope of indicators included in the study, whose expansion can enrich the analysis. For example, this study focused on the “embodied” carbon emissions of the buildings, since databases and methods are needed to tackle the decarbonisation of building materials in the coming years [57]. Future studies could address a larger number of environmental impact categories as in the existing study [58], which focuses on comparing the correlation of the data from the environmental impact indicators from Environmental Product Declarations (EPDs). The analysis of the relationships between other environmental indicators could help to identify building materials that, in addition to facilitating climate change mitigation, also contribute towards mitigating other relevant environmental problems, such as pollution, habitat destruction, deforestation and loss of biodiversity.

5.3. Potential of the methodology and future developments

5.3.1. Contribution to sustainable design

This research demonstrates the potential of the methodology to better understand the growth logic of the different BCCA items concerning the three dimensions of sustainability. The results could benefit scientific knowledge and engineering practice by identifying, for example, building elements, techniques, and materials in any database to be used in projects that reduce the social, economic, and environmental impact of the life cycle of buildings. Therefore, based on the types of correlations obtained, sustainability strategies could be implemented. For example, four scenarios could be assessed (see Fig. 4), and if it is assumed that a high value in the indicators of each dimension would lead to a negative impact and vice versa, then impact reduction could be achieved in each scenario as follows.

- Scenario 1 (S1). Group of alternative items that present a strong positive correlation in all correlation coefficients. This is a favourable scenario for the application of similar strategies to reduce the impacts since the selection of items with the lowest values in the results of one dimension will lead to lower values in the results of the other two dimensions. Likewise, an increase in one dimension would negatively influence the other two dimensions.
- Scenario 2 (S2). Group of alternative items that present a strong negative correlation in all correlation coefficients. This is a scenario where the same strategy can affect the results in a different way, since one item with lower values in one dimension can have the highest values in another dimension or vice versa. In this case, it could be decided which dimension is the one to be prioritised, in the knowledge that this action would be detrimental to the other two dimensions.
- Scenario 3 (S3). Group of alternative items that present a weak or null correlation in all correlation coefficients. In this case there is no apparent relationship between the dimensions. This scenario requires an item-by-item detailed study to analyse the incidence of the results of each dimension and the growth trend, even if it is weak, with respect the others. In this scenario, sustainable design measures that may affect one dimension to a greater extent do not necessarily affect the remaining dimensions.
- Scenario 4 (S4). Group of alternative elements that present a strong correlation (positive or negative) in one dimension and a strong (with another sign), weak, or non-existent correlation in the remaining dimensions. This scenario will require a case-by-case study. Items with lower values in one of the dimensions could have higher values in the second directly correlated

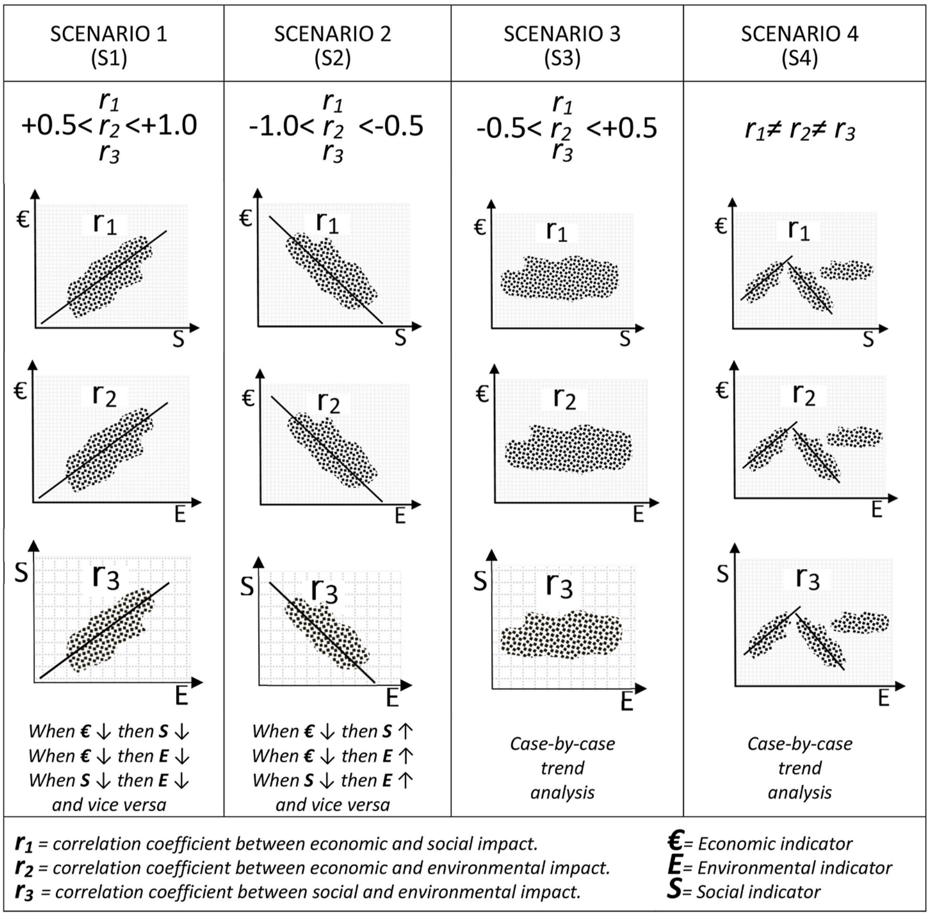


Fig. 4. Scenarios (S) to consider in the development of sustainability strategies in design.

dimension. Furthermore, the behaviour of the third dimension should be verified, in case there is no clear group tendency detected.

In the case of the analysed items of the TBL-BIM database, three S1 (e.g., m^3 Foundations), no S2, six S3 (e.g., m Masonry), and thirteen S4 (e.g., m^3 structure) scenarios were identified, as shown in Table 2.

In S1, the selection of items with lower results in one dimension will imply lower results in the other two dimensions. The significance of this fact lies in the identification of construction techniques with the lowest impact in all three dimensions. For example, the study reveals that in the case of earthwork (m^3) the transportation of soil by mechanical means would reduce not only the economic cost, but also the environmental and social impact (Figure S1.19 in supplementary data). Alternatively, in the foundations (m^3), the most economical types of concrete are those that also have the lowest environmental and social impacts (Figure S1.20).

In S3, moderate and non-existent correlations were obtained, such as in BCCA items for floor finishing, ceilings, carpentry, and exterior fittings (Figures S13 to S15, and S18). In these cases, the relationship between the variables tends to be horizontal since there are no peaks and troughs that follow a trend. Hence, the variability is related to the distribution curve of the data, which was neither clearly incremental nor decreasing in any dimension. However, the lack of clear correlations indicates that the variability of the values is not very marked, which in turn shows that values behave similarly in the three dimensions. Other correlation distributions will merit a detailed study to obtain trends.

In S4, several sub-scenarios can occur. As in S3, it will be necessary to carry out an in-depth study of two-by-two dimensions. For example, the coexistence of strongly positive correlations between two dimensions and strongly negative correlations between two others could limit the application of strategies aimed at reducing the values of all three dimensions at the same time. In such cases, one dimension will be favoured over another. This is the case of foundations (kg) (Fig. 2a), where the items with the highest values in the economic and environmental dimensions have the lowest values in the social dimension, and vice versa.

A last illustrative example of S4 is given by the structure chapter (m^3). In that case, the obtained values indicate that the growing tendency for economic and social dimension is less accelerated than for the environmental dimension, as shown in supplementary data (Figure S1.21). This means that the more materials are used in the structure, the greater the increase in the trends of environmental impacts will be in several items, greater than those of the economic and social dimension. This knowledge extracted from the analysis could prove useful in decision-making in sustainable design in professional practice. For example, if the environmental criterion prevails, then the optimisation of the building elements of the Structure chapter (beams, pillars, walls), with functional unit "m^3", will have a major contribution in the reduction of the environmental impact.

5.3.2. Future work

This study has provided interesting findings on how to improve and optimise the implementation of Data Mining and KDD techniques in this field. The results show that the way the values of the different dimensions grow is not directly related to the functional unit or chapter under consideration and should therefore be analysed per element on a case-by-case basis. This would lead to further studies including fewer values but organised in terms of similar building elements to analyse the data more specifically.

Future work should focus on such developments, which may also be useful for designers and architects to use in the creation of LCSA data adapted to the BIM workflow [15], including information on environmental, economic, and social dimensions throughout the life cycle of the building. Further indicators should be included in the TBL-database as should other LCSA modules.

The method also contributes to the field of data interpretation and visualisation in LCSA, thereby complementing existing frameworks and formats described in other studies such as [59]. In this vein, future work can also focus on integrating other parameters (such as geometry and dimensions of the building elements, type of materials, and bill of material quantities) into the analysis, which can help the designer to identify their influence on the LCSA data (regarding the building elements) and the LCSA results. The more optimised the model is in terms of the use of materials and dimensions of the elements that comprise the building, the better the results of the triple assessment become. Moreover, unlike other databases, the systematised integration of the database into the BIM workflow would enable the analysis of BIM models to be automated.

6. Conclusions

The literature review provides evidence that the analysis of three-dimensional data based on the building construction and building process to support the LCSA implementation in BIM has scarcely been addressed. This gap in the literature generates opportunities in the development of data sources. This study focuses on filling these gaps, by enriching the existing BIM-TBL database and enabling a comprehensive analysis of the building data to support the LCSA implementation. The complexity of the LCSA encompasses not only the collection and harmonisation of data but also the correlation between different and occasionally opposite values. This paper shows the application of KDD technology and specifically targets Data Mining to improve the use of information regarding the building and its environmental, economic, and social impacts. The aim is to better understand possible strategies for the design of buildings of a more sustainable nature. The first main contribution of the paper involves the enrichment of the BIM-TBL database. This enables the complexity of the data collection to be understood regarding the application of the method on the BCCA items as well as that of its processing on a large number of building elements. This research constitutes the first application of the KDD process to the data sources for triple bottom line sustainability assessment for the implementation of LCSA into buildings in BIM. This has enabled the RQs to be answered, and hence the following conclusions can be drawn.

- Various levels of correlation were found, depending on the type of element, functional unit, and its material, thereby answering RQ1.

- The study most frequently found correlations between the economic and social (r_1) dimensions (68 %), followed by correlations between the economic and environmental (r_2) dimensions (32 %), and by the non-existent or negative correlations between the social and environmental (r_3) dimensions (18 %), thereby answering RQ2.
- Findings from the correlation analysis of environmental, economic, and social dimensions reveal their influence on the type of building system and building element.
- Accordingly, based on these correlations, four scenarios were identified to aid in sustainable decision-making. In general, the growth patterns of the different dimensions are similar, which could favour the triple assessment, since contradictory behaviour patterns could complicate the application of the LCSA on the BCCA items.
- Future studies should enrich the database with other environmental, economic, and social impact categories and other life cycle phases. In this vein, considering the triple-dimension assessment, special attention should be paid to the analysis and weighting of these three dimensions. Further to this type of correlation analysis, the triple-dimension assessment should always include a critical evaluation of the weighting of each dimension, since the least environmentally damaging solutions may well not be the most economically viable. As a result of this type of analysis, the pathway towards the reduction of impacts should be prioritised, while being more affordable and beneficial to society.

Funding

This publication is part of the following projects: Grant TED2021-129542B–I00, funded by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR” and Grant PID2022-137650OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by FEDER, EU.

CRediT authorship contribution statement

Carmen Llatas: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization. **Bernardette Soust-Verdaguer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Luis Castro Torres:** Visualization, Validation, Software, Investigation, Formal analysis. **Daniel Cagigas:** Visualization, Validation, Supervision, Software, Data curation.

Declaration of competing interest

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

Data availability

The data are provided in the supplementary data file.

Acknowledgement

The authors thank the participants in the above projects for their direct and indirect contributions to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jobe.2024.109546.

References

[1] International Energy Agency (IEA), Emissions – Global Energy & CO2 Status Report 2019 – Analysis - IEA, 2019. https://www.iea.org/reports/global-energy-co2-status-report-2019. (Accessed 1 April 2024).

[2] UN - United Nations, The sustainable development goals Report 2019, https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf, 2019. (Accessed 1 April 2024).

[3] European Commission, European parliament, the European green deal, https://doi.org/10.1017/CBO9781107415324.004, 2019. (Accessed 1 April 2024).

[4] European Parliament, Directive of the European Parliament and of the Council on the energy performance of buildings (recast) COM/2021/802 final, https://doi.org/10.1108/ss.2005.11023aab.001, 2021. (Accessed 1 April 2024).

[5] ISO, ISO 14040:2006 Environmental Management — Life Cycle Assessment — Principles and Framework, 2006. https://www.iso.org/standard/37456.html. (Accessed 1 April 2024).

[6] M. Almokdad, R. Zentar, Flash-calcined sediments versus raw sediments: a comparative life cycle assessment of SCMs, Construct. Build. Mater. (2024) 134550, https://doi.org/10.1016/j.conbuildmat.2023.134550.

[7] Z. Mkahal, W. Maherzi, Y. Mamindy-Pajany, B. Bouzar, N.-E. Abriak, Development of a low-carbon binder based on raw, ground, and carbonated waste paper fly ash, Sustain. Mater. Technol. 36 (2023) e00650, https://doi.org/10.1016/j.susmat.2023.e00650.

[8] M. Bahramian, K. Yetilmezsoy, Life cycle assessment of the building industry: an overview of two decades of research (1995–2018), Energy Build. 219 (2020) 109917, https://doi.org/10.1016/j.enbuild.2020.109917.

[9] H. Lei, L. Li, W. Yang, Y. Bian, C.-Q. Li, An analytical review on application of life cycle assessment in circular economy for built environment, J. Build. Eng. 44 (2021) 103374, https://doi.org/10.1016/j.jobe.2021.103374.

[10] H. Feng, J. Zhao, H. Zhang, S. Zhu, D. Li, N. Thurairajah, Uncertainties in whole-building life cycle assessment: a systematic review, J. Build. Eng. 50 (2022) 104191, https://doi.org/10.1016/J.JOBE.2022.104191.

[11] S. Valdivia, C. Ugaya, G. Sonnemann, J. Hildenbrand (Eds.), Towards a Life Cycle Sustainability Assessment. Making Informed Choices on Products, Paris, 2011. https://www.unep.org/resources/report/towards-life-cycle-sustainability-assessment-making-informed-choices-products. (Accessed 1 April 2024).

[12] S. Somanath, A. Hollberg, S. Beemsterboer, H. Wallbaum, The relation between social life cycle assessment and green building certification systems, in: 7th Soc. LCA Conf., 2020, pp. 198–201. https://www.fruitrop.com/en/media/Publications/FruiTrop-Thema/Social-LCA-volume-5-7th-SocSem#book. (Accessed 1 April 2024).

[13] J.G. Backes, M. Traverso, Application of life cycle sustainability assessment in the construction sector: a systematic literature review, Processes 9 (7) (2021) 1248, https://doi.org/10.3390/pr9071248.

[14] J.G. Backes, M. Traverso, Life cycle sustainability assessment as a metrics towards SDGs agenda 2030, Curr. Opin. Green Sustainable Chem. 38 (2022) 100683, https://doi.org/10.1016/j.cogsc.2022.100683.

[15] C. Llatas, B. Soust-Verdaguer, A. Passer, Implementing life cycle sustainability assessment during design stages in building information modelling: from systematic literature review to a methodological approach, Buiding Environ. 182 (2020) 107164, https://doi.org/10.1016/j.buildenv.2020.107164.

[16] V.G. Larsen, N. Tollin, P.A.M.B. Sattrup, H. Tine, What are the challenges in assessing circular economy for the built environment? A literature review on integrating LCA, LCC and S-LCA in life cycle sustainability assessment, LCSA, J. Build. Eng. 50 (2022) 104203, https://doi.org/10.1016/j.jobe.2022.104203.

[17] C. Boje, Á.J.H. Menacho, A. Marvuglia, E. Benetto, S. Kubicki, T. Schaubroeck, T. Navarrete Gutiérrez, A framework using BIM and digital twins in facilitating LCSA for buildings, J. Build. Eng. 76 (2023) 107232, https://doi.org/10.1016/j.jobe.2023.107232.

[18] N. Kohler, S. Moffatt, Life-cycle analysis of the built environment, UNEP Ind. Environ. 26 (2) (2003) 17–21. https://www.researchgate.net/publication/279711810_Life-cycle_analysis_of_the_built_environment#fullTextFileContent.

[19] Directive 2014/24/EU. Official Journal of the European Union, 2014. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0024&from=es. (Accessed 1 April 2024).

[20] B. Soust-Verdaguer, I. Bernardino Galeana, C. Llatas, M.V. Montes, E. Hoxha, A. Passer, How to conduct consistent environmental, economic, and social assessment during the building design process. A BIM-based Life Cycle Sustainability Assessment method, J. Build. Eng. 45 (2021) 103516, https://doi.org/10.1016/J.JOBE.2021.103516.

[21] Itec, BEDEC, 2020. https://metabase.itec.cat/. (Accessed 1 April 2024).

[22] C. Llatas, B. Soust-Verdaguer, A. Hollberg, E. Palumbo, R. Quiñones, BIM-based LCSA application in the early design stages using IFC, Autom. ConStruct. 138 (2022) 104259, https://doi.org/10.1016/J.AUTCON.2022.104259.

[23] ISO, ISO 21931-1:2022 Sustainability in building construction -- Framework for methods of assessment of the environmental, economic and social performance of construction works -- Part 1: buildings, https://www.iso.org/standard/71183.html#:~:text=This%20document%20provides%20a%20general,the%20sustainability%20assessment%20of%20buildings, 2022. (Accessed 1 April 2024).

[24] Andalusian Government, BCCA, Base de Costes de la Construcción de Andalucía. Clasificación Sistemática de Precios Básicos. Auxiliares Y Unitarios, 2023. https://www.juntadeandalucia.es/organismos/fomentoarticulaciondelterritorioyvivienda/areas/vivienda-rehabilitacion/planes-instrumentos/paginas/vivienda-bcca.html. (Accessed 1 April 2024).

[25] J.-R. Lee, S.-L. Hsueh, H.-P. Tseng, Utilizing data mining to discover knowledge in construction enterprise performance records, J. Civ. Eng. Manag. 14 (2008) 79–84, https://doi.org/10.3846/1392-3730.2008.14.2.

[26] H.-C. Hsu, S. Chang, C.-C. Chen, I.-C. Wu, Knowledge-based system for resolving design clashes in building information models, Autom. ConStruct. 110 (2020) 103001, https://doi.org/10.1016/j.autcon.2019.103001.

[27] E. Petrova, K.S. Pieter Pauwels, R. Lund Jensen, Search of sustainable design patterns: combining data mining and semantic data modelling on disparate building data, in: Adv. Informatics Comput. Civ. Constr. Eng. Proc. 35th CIB W78 2018 Conf. IT Des. Constr. Manag., 2019, pp. 19–26, https://doi.org/10.1007/978-3-030-00220-6_3.

[28] S.A. Cype Ingenieros, CYPE. Generador de Precios, 2019. http://www.maroc.prix-construction.info/. (Accessed 1 April 2024).

[29] K. Figueiredo, R. Pierott, A.W.A. Hammad, A. Haddad, Sustainable material choice for construction projects: a Life Cycle Sustainability Assessment framework based on BIM and Fuzzy-AHP, Build. Environ. 196 (2021) 107805, https://doi.org/10.1016/j.buildenv.2021.107805.

[30] Z.M. Salehabadi, R. Ruparathna, User-centric sustainability assessment of single family detached homes (SFDH): a BIM-based methodological framework, J. Build. Eng. 50 (2022) 104139, https://doi.org/10.1016/j.jobe.2022.104139.

[31] R. Frischknecht, N. Jungbluth, H.-J. Althaus, G. Doka, R. Dones, T. Heck, S. Hellweg, R. Hischier, T. Nemecek, G. Rebitzer, M. Spielmann, Ecoinvent Version 2. The ecoinvent database: overview and methodological framework, Int. J. Life Cycle Assess. (2005) 3–9, https://doi.org/10.1065/lca2004.10.181.1.

[32] RSmeans, RSMeans data: construction cost estimating software, https://www.rsmeans.com/, 2019.

[33] N. Greg, C. Benoit Norris, N. Chahande, Social Hotspots, http://www.socialhotspot.org/, 2023. (Accessed 1 April 2024).

[34] B. Soust-Verdaguer, T. Potrč Obrecht, N. Alaux, E. Hoxha, M. R M Saade, M. Röck, A. Garcia-Martinez, C. Llatas, J. Gómez de Cózar, A. Passer, Using systematic building decomposition for implementing LCA: the results of a comparative analysis as part of IEA EBC Annex 72, J. Clean. Prod. 284 (2023) 135422, https://doi.org/10.1016/j.jclepro.2022.135422.

[35] B. Soust-Verdaguer, J.A. Gutiérrez Moreno, C. Llatas, Development of a plug-in to support sustainability assessment in the decision-making of a building envelope refurbishment, Buildings 13 (6) (2023) 1472, https://doi.org/10.3390/buildings13061472.

[36] H. Yan, N. Yang, Y. Peng, Y. Ren, Data mining in the construction industry: present status, opportunities, and future trends, Autom. ConStruct. 119 (2020) 103331, https://doi.org/10.1016/j.autcon.2020.103331.

[37] J. Zhao, Refined node data mining platform in the construction phase of building installation engineering based on BIM technology, in: 2022 6th Int. Conf. Intell. Comput. Control Syst., 2022, pp. 1717–1720, https://doi.org/10.1109/ICICCS53718.2022.9788279.

[38] E.A. Petrova, AI for BIM-Based Sustainable Building. Design Integrating Knowledge Discovery and Semantic Data Modelling for Evidence-Based Design Decision Support, Aalburg University, 2019. https://vbn.aau.dk/en/publications/ai-for-bim-based-sustainable-building-design-integrating-knowledg. (Accessed 1 April 2024).

[39] X. Pan, B. Zhong, X. Wang, R. Xiang, Text mining-based patent analysis of BIM application in construction, J. Civ. Eng. Manag. 27 (2021) 303–315, https://doi.org/10.3846/jcem.2021.14907.

[40] Y. Pan, L. Zhang, BIM log mining: learning and predicting design commands, Autom. ConStruct. 112 (2020) 103107, https://doi.org/10.1016/j.autcon.2020.103107.

[41] Y. Pan, L. Zhang, Modeling and analyzing dynamic social networks for behavioral pattern discovery in collaborative design, Adv. Eng. Inf. 54 (2022) 101758, https://doi.org/10.1016/j.aei.2022.101758.

[42] Y. He, Y. Chu, Y. Song, M. Liu, S. Shi, X. Chen, Analysis of design strategy of energy efficient buildings based on databases by using data mining and statistical metrics approach, Energy Build. 258 (2022) 111811, https://doi.org/10.1016/j.enbuild.2021.111811.

[43] Y. Qian, J. Leng, CIM-based modeling and simulating technology roadmap for maintaining and managing Chinese rural traditional residential dwellings, J. Build. Eng. 44 (2021) 103248, https://doi.org/10.1016/j.jobe.2021.103248.

[44] L. Soibelman, H. Kim, Data preparation process for construction knowledge generation through knowledge discovery in databases, J. Comput. Civ. Eng. 16 (2002) 39–48, https://doi.org/10.1061/(ASCE)0887-3801(2002)16:1(39).

[45] M.M. Ghazal, A. Hammad, Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects, Int. J. Constr. Manag. 22 (2022) 1632–1646, https://doi.org/10.1080/15623599.2020.1738205.

[46] A. Hammad, S. AbouRizk, M. Y., Application of KDD techniques to extract useful knowledge from labor resources data in industrial construction projects, J. Manag. Eng. 30 (2014) 5014011, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000280.

[47] B. Soust-Verdaguer, J.A. Gutiérrez Moreno, C. Llatas, Utilization of an automatic tool for building material selection by integrating life cycle sustainability assessment in the early design stages in BIM, Sustain. Times 15 (3) (2023) 2274, https://doi.org/10.3390/su15032274.

[48] C. Benoît Norris, M. Traverso, S. Neugebauer, E. Ekener, T. Schaubroeck, S. Russo Garrido, M. Berger, S. Valdivia, A. Lehmann, M. Finkbeiner, G. Arcese (Eds.), UNEP. Guidelines for Social Life Cycle Assessment of Products and Organizations, 2020. https://www.lifecycleinitiative.org/wp-content/uploads/2020/12/Guidelines-for-Social-Life-Cycle-Assessment-of-Products-and-Organizations-2020-sml.pdf. (Accessed 1 April 2024).

[49] G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, B. Weidema, Ecoinvent Version 3. The ecoinvent database version 3 (part I): overview and methodology, Int. J. Life Cycle Assess. 21 (2016) 1218–1230, https://doi.org/10.1007/s11367-016-1087-8.

[50] Microsoft, microsoft access, (n.d.), https://www.microsoft.com/en-us/microsoft-365/access. (Accessed 1 April 2024).

[51] D. Freedman, R. Pisani, R. Purves, Statistics: fourth international student edition, 0393930432, 9780393930436, W. W. Norton & Company, 2007. https://books.google.es/books/about/Statistics.html?id=vWeQDAAAQBAJ&redir_esc=y. (Accessed 1 April 2024).

[52] Y. Dodge, Kendall Rank Correlation Coefficient BT - the Concise Encyclopedia of Statistics, Springer New York, New York, NY, 2008, pp. 278–281, https://doi.org/10.1007/978-0-387-32833-1_211.

[53] T.W. MacFarland, J.M. Yates, in: T.W. MacFarland, J.M. Yates (Eds.), Spearman's Rank-Difference Coefficient of Correlation BT - Introduction to Nonparametric Statistics for the Biological Sciences Using R, Springer International Publishing, Cham, 2016, pp. 249–297, https://doi.org/10.1007/978-3-319-30634-6_8.

[54] A.M. Jarkas, C. Bitar, Factors affecting construction labor productivity in Kuwait, J. Construct. Eng. Manag. 138 (2011) 811–820, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000501.

[55] Y. Martinez-Martin, F. Gonzalez-Vidosa, A. Hospitaler, V. Yepes, Multi-objective optimization design of bridge piers with hybrid heuristic algorithms, J. Zhejiang Univ. - Sci. 13 (2012) 420–432, https://doi.org/10.1631/jzus.A1100304.

[56] S. Eleftheriadis, P. Duffour, P. Greening, J. James, B. Stephenson, D. Mumovic, Investigating relationships between cost and CO_2 emissions in reinforced concrete structures using a BIM-based design optimisation approach, Energy Build. 166 (2018) 330–346, https://doi.org/10.1016/j.enbuild.2018.01.059.

[57] United Nations Environment Programme (UNEP), & Yale Center for Ecosystems + Architecture, Building materials and the climate: constructing a new future, https://wedocs.unep.org/20.500.11822/43293, 2023.

[58] V.E.M. Cardoso, L. Sanhudo, J.D. Silvestre, M. Almeida, A. Aguiar Costa, Challenges in the harmonisation and digitalisation of Environmental Product Declarations for construction products in the European context, Int. J. Life Cycle Assess. 35 (2024), https://doi.org/10.1007/s11367-024-02279-w.

[59] J.G. Backes, L.S. Steinberg, A. Weniger, M. Traverso, Visualization and interpretation of life cycle sustainability assessment—existing tools and future development, Sustain. Times 15 (2023) 10658, https://doi.org/10.3390/su151310658.