Preprints.org

Article

Closing Data Gaps with Emission Factors: A General Method Applied to EPDs in Brightway

Vanessa Schindler [*] , Moritz Kohlhase , Heidi Hottenroth , Ingela Tietze

Article

Closing Data Gaps with Emission Factors: A General Method Applied to EPDs in Brightway

Vanessa Schindler [1], Moritz Kohlhase [2], Heidi Hottenroth [3] and Ingela Tietze[3]

[1] Institute for Resource Efficiency and Energy Strategies, 76131 Karlsruhe, Germany

[2] WoodenValley gGmbH, Quellenstraße 7a, 70376 Stuttgart, Germany

[3] Pforzheim University, Institute for Industrial Ecology, Tiefenbronner Straße 65, 75175 Pforzheim, Germany

* Correspondence: heidi.hottenroth@hs-pforzheim.de, +49 7231 286403

Abstract: Purpose Data gaps decrease the accuracy and reliability of Life Cycle Assessment (LCA). Although reported environmental data from companies could fill these gaps, they lack the process-based supply chain information required by common Life Cycle Inventory (LCI) databases. Instead of trying to integrate such process-based information, emission factors from environmental data can be directly integrated in LCA. This approach, however, requires in-depth understanding of LCA software and intensive maintenance, reducing user-friendliness. **Methods** This paper presents a user-friendly method that integrates aggregated emission factors from environmental product declarations (EPD) into LCI. The method includes reverse calculation and proxy selection. Thus, proxies represent ecosphere flows to allow for inventory result calculation in life cycle impact assessment (LCIA). The method generates LCI databases from emission factors with corresponding proxies, which provide these emission factors as LCA results within common LCA programs. **Results and discussion** The approach is demonstrated by integrating the ökobaudat (OBD) database – containing EPDs - into the LCA software Brightway. An example of LCA calculations with data from conventional and EPD databases using Brightway is provided. This example illustrates how data gaps can be effectively filled in a user-friendly way by simultaneously using data with and without supply chain information. **Conclusions** The feasibility of incorporating data without supply chain information in conventional LCA software is demonstrated. The developed method increases user-friendliness, as only one LCA modelling and calculation is required when using data with and without process-based information simultaneously. Additionally, the presented method is not restricted to a specific set of data sources, emission factors, or LCA programs and is generally applicable.

Keywords: life cycle assessment; life cycle inventory; data gaps; environmental product declaration; emission factor; sustainability; buildings; brightway; ökobaudat

1. Introduction

Is it better to assume zero or use imperfect data to fill data gaps in Life Cycle Assessment (LCA)? Data gaps are a common issue in LCA and must be addressed for a meaningful interpretation of the study to support decision-making. Del Rosario et al. (2021) posit that data gaps have a substantial impact on the interpretation of LCA studies. They conduct two similar LCA studies of a sustainable building with and without significant data gaps. The first LCA is based on EPDs and has data gaps in the transportation scenarios, whereas the second one uses ecoinvent data to fill these gaps. Above all, filling data gaps in the second LCA has led to a significant increase in all results in most indicators compared to the first LCA. The data used to fill the data gaps, representing the transportation, significantly contributes to the results and assuming them to be zero leads to biased results. If relevant data in LCA is missing, there is a growing risk that decisions cannot be made, or worse, they are made incorrectly due to biased results.

Specifically, data gaps in primary and secondary data can determine the environmental impact of products or services. To conduct an LCA of a product or service, its upstream and downstream process structure must not only be known but also be quantifiable to determine the environmental influences. To quantify company-specific foreground processes, it is necessary to obtain primary data. At a company level, it is straightforward for the LCA modeller to obtain primary information from internal sources regarding direct emissions and purchased energy (Baehr et al. 2024). This primary data is represented in the foreground system of an LCA whereas secondary data is commonly used for the background system and depicts average processes (Hauschild et al. 2017, p. 80). In the background system, Life Cycle Inventory (LCI) databases with secondary data are employed for modelling the process structure of the upstream processes of purchased goods or services. However, the background system often lacks environmental information on specific processes although the LCI databases are continuously being expanded (ecoinvent 2023). Hence, it is essential to have a background system that well represents upstream processes as well as a consistent foreground system with primary data. The absence of data gaps in the systems pertaining to the foreground and background is imperative for the evaluation of   the environmental impact of goods or services.

It is necessary to implement strategies to fill data gaps to obtain consistent product systems in the LCA model. These strategies include proxy selection and data creation (Zargar et al. 2022). Proxy selection, i.e. selecting a similar data set as a proxy, inevitably results in a reduction in data quality. Data creation through empirical evaluations or the result analysis from other models is highly elaborate. Neither proxy selection nor data creation are ideal solutions to fill data gaps. It is therefore necessary to identify simpler ways to fill data gaps for meaningful LCA results.

1.1. Existing Environmental Data Sources

However, environmental data can already fill data gaps and be integrated into LCA. A large number of organizations already provide specific environmental data due to mandatory or volutary reporting acitivities. In accordance with the EU Non-Financial Reporting Directive or the Corporate Sustainability Reporting Directive (2022), sustainability reporting is mandatory for several companies. Alternatively, organizations may choose to report voluntarily in accordance with ISO Net Zero Standards (2023), the Greenhouse Gas Protocol (World Resources Institute and World Business Council et al. 2004) or the corporate standard of the Science Based Targets initiative (SBTi) (2024). All these reporting activities generate specific environmental data. Consequently, these reported environmental data can be used as input data as a source to fill data gaps in other LCA studies. (Baehr et al. 2024)

Standardized environmental data is available as a reliable data source. Established standards already exist to facilitate comparison of reported environmental data within a company but also along the supply chain (Fritz et al. 2017; de Villiers et al. 2022). Formats for the standardized reporting of the life cycle environmental performance of products are not only environmental product declarations (EPDs) adhere to the ISO EN 14025 but also the more specific standard for EPDs of the building sector, outlined in EN 15804 (Deutsches Institut für Normung 2022). Information regarding EPD data is available from, e.g. Institut Bauen und Umwelt e.V. (2024), and the International EPD® System, where approximately 23,000 validated EPDs are available as of 2024. Other standardized schemes for the reporting of the life cycle environmental performance of products include the product carbon footprint in accordance with ISO 14067 or the Greenhouse Gas Protocol Standard as well as the carbon footprint in accordance with regional standards such as the European Commission's recommendation (2021). All these reporting standards for life cycle environmental data follow rules, which makes them a reliable data source to fill data gaps.

However, this standardized environmental data cannot be integrated into common LCI databases due to their aggregation level. They are aggregated concerning life cycle phases and unit processes on the one hand and characterized impacts on the other hand. Intermediates or elementary flows which could be subject to inventory or impact indicator calculation are not available. Another

limitation for use in LCA is the availabiltiy of only few impact categories, usually only global warming potential emission factor, e.g., from the Greenhouse Gas Protocol (2016) the Intergovernmental Panel on Climate Change (IPCC) (2021) or the German Umweltbundesamt (2022). EPDs contain emission factors for 19 defined impact categories of the impact assessment method Environmental Footprint (EF) (European Commission 2023) in accordance with the EN 15804 (Deutsches Institut für Normung 2022). Each emission factor for one of these impact categories, e.g. the carbon footprint, can be interpreted as a limited LCA and is one characterization result within an LCA (Meinrenken et al. 2020). The underlying data used to calculate the inventory results comes usually from to process-based LCI databases, which represent the most disaggregated level of environmental data and cover information on the entire supply chain. For instance, emission factors can be further aggregated e.g. in EPDs and disaggregated based on its supply chain information in LCA. When it comes to the transfer of emission factors for intermediate products to the subsequent company in the supply chain, the transferred information is limited to aggregated environmental information without process information or any other methodological description (Schmidt 2011). Although emission factors are generally derived from process-based LCI databases and therefore contain supply chain information, this information is aggregated into a single numerical value. Hence, the underlying process chain becomes inaccessible when this aggregation occurs. This limits the integration of emission factors into LCI databases, which rely on detailed process-based information to calculate inventory results. Although emission factors are derived from LCI databases, their underlying process structure is lost during aggregation, preventing their reintegration into such databases.

Building on these points, it becomes evident that environmental data might fill data gaps in LCA. To fill data gaps on a large scale, standardized environmental data such as EPD is necessary, but currently cannot be integrated into common LCI databases due to inaccessible process-based information.

1.2. Integrate Emission Factors from EPDs in LCA

The central research question concerns how emission factors from standardized EPDs can be systematically integrated into LCA methodologies. Although environmental data sources—such as emission factors provided in EPDs or company-reported environmental metrics—are increasingly available, they typically lack detailed, process-based information. As a result, the data structures used in EPDs differ fundamentally from those employed in conventional LCI databases, which are grounded in process-based modeling. This structural incompatibility presents a barrier to directly integrating EPD-derived data into standard LCI databases. Nevertheless, the use of emission factors as an alternative or supplementary LCA data source remains feasible, albeit methodologically challenging. Integrating such factors could help address existing data gaps in LCA. Therefore, it is essential to develop approaches that allow for the concurrent use of both process-based LCI databases and emission factor–based data sources within a unified LCA framework. Methods already exist which are using emission factors of EPDs as data sources for LCA together with LCI databases, which are summarized in the following.

The literature shows two approaches for the integration of EPDs in LCA software. A first approach is a reconstruction method presented by Zhang et al. (2024). They connect the EPD database KBOB, which has outdated background information, to the up-to-date ecoinvent database for an actual and harmonized LCI database. This integration process eliminates the problem of missing process-based background data and makes it straightforward to implement these EPDs in LCA. Abu-Ghaida et al. (2024) also use a reconstruction method to integrate EPD in LCA. In contrast to Zhang et al. (2024), the background information of their EPDs is unknown, but similarly, they link their EPDs to the ecoinvent database. In addition to the doubled work of the reconstruction process, which can be seen in both studies, it is unclear whether, in the study of Abu-Ghaida et al. (2024), the equivalent products from ecoinvent are an adequate representation of the EPDs. These studies present the method of reconstructing EPDs by using the ecoinvent database for LCA. A shortcoming

of this first straightforward suggestion in literature, the reconstruction of EPDs with LCI databases is the doubling of the work as EPDs already provide environmental data.

Second, creating new impact categories is another approach to display the EPD emission factor results in LCA. Emara and Ciroth (2014) create new impact categories to import an EPD database in the conventional LCA software openLCA and SimaPro. They add new impact categories to the LCA software which can deal with the available emission factors in the EPDs. These new impact categories sum up emission factors from EPDs but cannot deal with process-based LCI databases. These process-based LCI databases do not contain emission factors, but environmental flows which have different units than emission factors and therefore cannot be summed up. Within this method of new impact categories, EPDs and process-based data cannot be used in one LCA calculation simultaneously. In addition, the authors note that the contribution analysis shows few details in the form of the contribution of each EPD as background information is lacking. Until now, this is the most promising method, but the implementation of these new impact categories needs changes in the existing LCA program as the LCIA is not performed when calculating an LCA with EPDs, which is displayed in the middle part of Figure 1. Including new impact categories in LCA software is the most promising method to integrate emission factors into LCA but is highly complex as the software has to be understood in detail for implementation.

Previous approaches have weaknesses and are not sufficient to make available emission factors in LCA to fill data gaps. Emission factors in EPDs in the absence of process-based data of the supply chain can be made available for LCA. It is not straightforward to combine emission factors from EPD together with LCI databases in LCA calculations as the lack of background supply chain information hinders the integration of emission factors from EPD as an LCI database in LCA software. The existing methods that utilise both, emission factors from EPD and LCI databases as a data source in LCA, have several disadvantages. The creation of new impact categories is maintenance-intensive and complex to implement and the generation of two results due to two calculations of the process-based and the non-process-based database is inconvenient. However, this integration would offer significant benefits, but previous approaches are not sufficient to integrate emission factors in LCA as they are neither user- nor developer-friendly.

The most advanced previous study lacks user-friendliness and needs two calculations when conducting an LCA with data from process-based databases and emission factors obtained from EPD. This study aims to develop a user-friendly approach for importing an EPD database into LCA programs to be used simultaneously with LCI databases in the same LCA. The methodology employed here is analogous to that described by Emara and Ciroth (2014) but shall avoid the presentation of two results stemming from the data structure of process-based LCI databases and EPD emission factors within a single LCA calculation.

Our approach is integrating the OBD database as an exemplary EPD database in the LCA software Brightway using reverse calculation and impact category proxies. Unlike Emara and Ciroth (2014), we do not create new impact categories but rather employ reverse calculation and impact category proxies so that emission factors can be processed by the existing impact categories. By avoiding the creation of new impact categories, we ensure that only one result is presented per LCA calculation.

We do this by describing the method of integrating emission factors from the ökobaudat (OBD) [1] database (Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen 2024), an EPD database, in Brightway and the Activity-Browser (Steubing et al. 2020). The description of the method is therefore divided into sections to identify relevant differences in the structure of EPD and process-based databases, how results are calculated and displayed by the LCA software Brightway and the Activity-Browser and the approach of reverse calculation and impact category proxies. Based on this, the results present the chosen mathematical representations applied to each impact category, along

---

[1]This database is provided by the German Federal Ministry for Housing, Urban Development and Building at no cost.

https://www.oekobaudat.de/en/service/downloads.html

with the technical implementation in Python and an illustrative LCA calculation. Furthermore, the discussion section identifies the limitations of the methodology. The paper ends with concluding remarks and recommendations for future research.

2. Methods: Integrating Emission Factors in LCA Programs

Firstly, it is imperative to comprehend the structural intricacies of EPD datasets in contradistinction to process-based datasets. The subsequent section of the text provides a detailed description of how LCA programs operate when production LCA results from LCI databases. Subsequently, the method of reverse calculation is presented, which details the integration of emission factors into LCA programs.

2.1. Differences Between EPD and Process-Based Databases

An LCA must be conducted to create an EPD in accordance with EN 15804. Thereby specifications are given for the goal and scope definition that facilitate comparison of EPD results. In the LCA, processes must be categorised into predefined construction phases, also known as modules. Module A encompasses all materials and the construction phase. Module B addresses the use phase and module C the end-of-life. Finally, module D covers the benefits of recycling. Subsequently, the defined impact categories must be utilized to obtain results for each module. The EPD documentation lacks any process-related information, and thus each impact category reported in an EPD can be interpreted as an emission factor. (International Standardization Organization 2006a; Deutsches Institut für Normung 2022)

EPDs might be stored in databases, such as the OBD, which have even a more specific structure.

The OBD is incorporating EPD data in accordance with the EN 15804 standard for the construction sector. This database contains approximately 1,400 EPD datasets that can be grouped as follows:

- specific datasets for products of a certain manufacturer,
- average datasets of products provided by multiple companies or standing for multiple plants or products, representative datasets for products in certain countries, template datasets for products with unspecific information,
- and generic datasets created using secondary data such as literature.

The OBD contains products each of which is identified by a unique identifier, the UUID. Each product contains LCIA results for all impact categories in accordance with EN 15804. The LCIA results are provided for each product, broken down by the life-cycle stage module (A-D). Furthermore, it should be noted that for some products different scenarios per module exist. For example, a given product is associated with one scenario using recycled parts and another scenario without recycled parts. These different scenarios lead to different emission effects for the same product. Furthermore, the data is varying, such that some products are associated with several scenarios, while others have none. Additionally, the number of modules specified varies across products, with some modules being specified in aggregated form (e.g. 'A1-A3'). (Figl and Kusche 2023)

The OBD database currently contains results for two impact assessment methods., but EN 15804-A2 will replace EN 15804-A1 in the future; hence, this study will focus on A2. The EN 15804 standard mandates the use of the EF impact assessment method and therefore the EF 3.0 impact categories. After the EN 15804 standard, results from six out of the 19 environmental impact categories are optionalas displayed in Table 1. The additional environmental impact categories are ecotoxicity, human toxicity (carcinogenic and non-carcinogenic), ionising radiation, particulate matter formation and land use. Furthermore, the Global Warming Potential (GWP) 100, land use and land use change may be excluded if its contribution is less than 5 % of the total GWP100 across the declared modules, except module D. (Figl and Kusche 2023)

Table 1. Obligatory and additional impact categories after theccording to EN 15804.

Obligatory impact categories	Additional impact categories
acidification	ecotoxicity
climate change	human toxicity: carcinogenic
climate change, biogenic	human toxicity: non-carcinogenic
climate change, fossil	ionising radiation
climate change, land use and land use change	land use
energy resources: non-renewable	particulate matter formation
Eutrophication, freshwater	
Eutrophication, marine	
Eutrophication, terrestrial	
material resources, metals/minerals	
ozone depletion	
photochemical oxidant formation: human health	
water use	

In conclusion, the OBD represents an EPD database with emission factors. It may be considered an exemplar of how to implement its information, which is described in the method below.

Differing from EPD databases, process-based LCI databases are the conventional data source for LCA. Process-based databases are designed to map the technosphere to model the entirety of a product's process structure throughout its life cycle, incorporating manufacturing, use phase and end-of-life (Hauschild et al. 2017, p. 149). The life cycle of a product as modelled in the process-based database is comprised of unit processes, which represent the smallest element (Hauschild et al. 2017, p. 76). These unit processes are interconnected by flows, which can either be inputs or outputs (Steen et al. 1995). Unit processes may be grouped to form products and intermediate products, thereby representing the concept of the technosphere, which encompasses technical and economic structures (Hauschild et al. 2017, p. 78). The technosphere is linked to the ecosphere, representing the natural environment (Hauschild et al. 2017, p. 78). In contrast to process-based databases, OBD contains emission factors and lack information about the supply chain of products.

2.2. Operation of LCA Programs for LCA Result Generation from LCI Databases

LCA programs orient their calculation on existing data sources such as process-based databases, as well as on established standards, namely the ISO 14040 and ISO 14044 standards (International Standardization Organization 2006b). In the context of a conventional LCA performed in LCA software in adherence with the ISO standards, the sequence of calculating impacts is predefined for two out of four LCA phases: the LCI and the LCIA. The LCI comprises technosphere and ecosphere flows. Technosphere flows are interlinkages within the LCI database and ecosphere flows interact with the environment (Hauschild et al. 2017, p. 79). These interlinkages to the ecosphere permit the aggregation of relevant ecosphere flows for each impact category with the subsequent conversion of these flows using characterisation factors to yield LCIA results per impact category (Hauschild et al. 2017, p. 81-82). The characterisation factors are dependent upon the impact categories selected for the LCIA. This approach is displayed in the upper part of Figure 1.

A common LCA software is the open-source LCA software Brightway2 (Mutel 2017) and its graphical user interface, the Activity-Browser (Steubing et al. 2020), which are utilized to implement the method. The open-source software Brightway2 provides the advantage of facilitating a more detailed understanding of the software's functionality of the code. The code of the LCA software Brightway is written in Python and performs the LCI and the LCIA in a predefined sequence. The

impact category results of the LCIA are automatically visualized in the software. (Mutel 2017; Steubing et al. 2020)

So far, LCA programs such as Brightway and the Activity-Browser cannot process emission factors or EPDs. Brightway is capable of working with process-based LCI databases, such as the ecoinvent database. Each database is a set of activities and the ecosystem is represented by the biosphere3 database in the open-source Python code. Because of the full transparency of open-source code, this software is the base for a fully automated and programming-based methodology for integrating emission factors and EPD into LCA software.

2.3. Reverse Calculation and Impact Category Proxies to Integrate Emission Factors in LCA

The integration of emission factors into process-based databases is achieved through the utilisation of reverse calculation and the inverse of the characterisation factor. In contrast to previous methods, which use the creation of novel impact categories for the LCIA and provide two distinct outcomes, our method develops impact category proxies. The differences between the conventional LCA approach, previous methods to include emission factors in LCA and our method of impact category proxies are shown in Figure 1. As displayed in the upper part of Figure 1, the conventional LCA approach uses LCI databases with process-based information to identify the ecosphere flows. These ecosphere flows are then multiplied with characterization factors to receive LCIA results. In the middle part of Figure 1, the most promising, existing method applied by Emara and Ciroth (2014) is displayed. They use emission factors as data source and create new impact categories to the LCA software which can transition the available emission factors from the data source to the LCA results. These new impact categories employ the proceeding of summing up emission factors, which differs significantly from the multiplication with characterization factors of the conventional LCA approach. This is why Emara and Ciroth (2014) implement an additional plugin for the LCA program to generate LCA results from emission factors, which are presented as a second outcome in addition to the conventional LCA results. Within this existing method of new impact categories, emission factors and process-based data cannot be used in one LCA calculation simultaneously, which is not user-friendly. Our method of impact category proxies is shown in the lower part of Figure 1. This new method can process emission factors and process-based data simultaneously and displays one LCA result. To provide only one outcome per LCA calculation, the impact category proxies must fulfil specific requirements to not affect the LCIA calculation of the LCA software. Therefore, the impact category proxies are initiated at the level of the LCI to perform the LCIA at the subsequent step. The impact category proxies consist of one or multiple ecosphere flows per proxy, and are implemented into the LCI using reverse calculation. The proxies link the emission factor database in the LCI to the ecosphere in the absence of any supply chain information. This implementation of the proxies ensures the correctness of the contribution per emission factor to the LCA result during the performance of the LCIA. The LCA results from our method utilising impact category proxies and reverse calculation can be directly summed up to conventional ones by the LCA program due to the structure of the approach. A comparative overview of methods is provided in Figure 1.
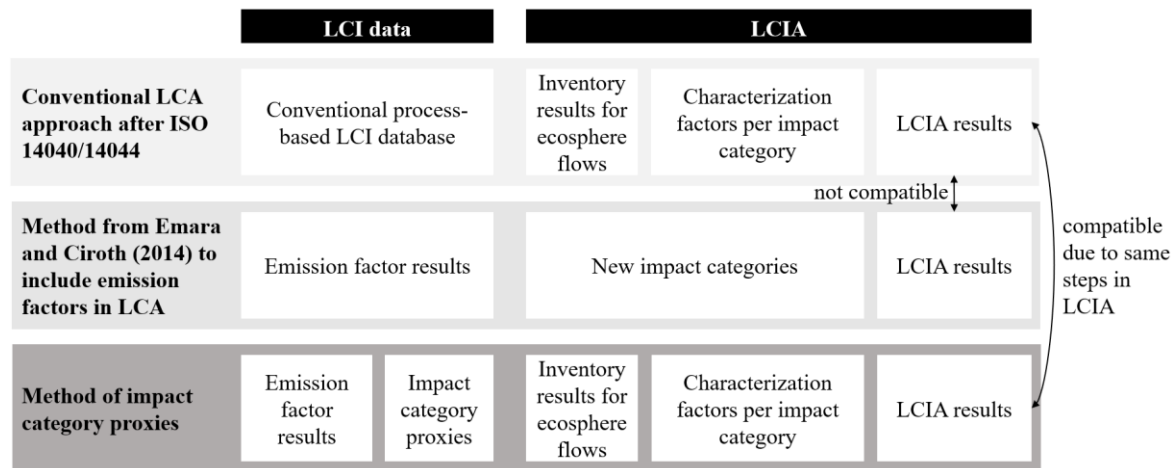
Figure 1. Overview of existing and new methodologies for emission factor integration in LCA in comparison to the conventional LCA approach.

The following section outlines the mathematically methodology for reverse calculation of emission factors for integration in LCI databases. The conventional approach serves as basis:

$E_k = \sum_i LCIA_{i,k} * m_i$ after (Heijungs 2022) (1)

E_k: Characterization result for each impact category k

$LCIA_{ik}$: Characterization factor for ecosphere flow i in category k

m_i: Inventory result for ecosphere flow i

The objective is to identify a possible conversion factor, called impact category proxy, for determining suitable inventory amounts m_i out of the emission factor results in the LCI, without influencing the LCIA. Consequently, our methodology exclusively affects m_i, whereas $LCIA_{ik}$ remains unaltered to maintain consistency with the calculation steps of a conventional LCA approach, as illustrated in Figure 1.

The LCA characterization results E_k, which represent the emission factors EF_k, are known. These emission factors need to be modified in a way, that the multiplication by the characterization factors $LCIA_{i,k}$ will result in the initial emission factors. Therefore, a reverse characterization factor has to be introduced. This means, the mathematical product of the impact category proxy and the characterisation factors $LCIA_{i,k}$ need to be one. The reverse characterization factor, referred to as the "$proxy_{k,i}$" for each impact category k referring to an ecosphere flow i, is defined as the inverse of the characterization factors $LCIA_{i,k}$.

$$LCIA_{i,k} * proxy_{k,i} = 1$$
$$proxy_{k,i} = \frac{1}{LCIA_{i,k}} \quad (2)$$

3. Results: Implementation of the OBD Database in Brightway2

For the implementation of EPDs, which in this example of the OBD are using EF v3.0 as impact assessment method, the approach has to ensure that the ecosphere flow i is exclusively representative of the impact category k. Therefore, it is essential that the ecosphere flow i is a unique flow (chapter 3.1) and does not appear in other impact categories. When the ecosphere flow i is a non-unique flow, it would result in the non-transparent influencing of results of other impact categories (chapters 3.2 and **Fehler! Verweisquelle konnte nicht gefunden werden.**). The emission factor results as well as the proxy database are stored within the LCA program in the same format as any process-based database as illustrated in Figure 1.

3.1. Finding Unique Flows in the Relevant Impact Categories to the Ecosphere

To identify unique ecosphere flows amongst all flows, a programming-based approach in Python has been employed. The code is accessible via GitHub[2] and is designed to quantify the number of duplicates within the list of all ecosphere flows of the relevant impact assessment method. This enables the selection of unique ecosphere flows within one impact assessment method for every impact category, where applicable.

In the event that at least one unique ecosphere flow, i_{unique}, is present within all relevant impact categories k in the LCA, the $proxy_{i_{unique}},$ can be expressed as follows for the impact categories k_{unique}:

$$proxy_{k_{unique}i_{unique}} = \frac{1}{LCIA_{i_{unique},k_{unique}}} \quad (3)$$

This $proxy_{i_{unique},k_{unique}}$ ensures that the ecosphere flow i is exclusively representative of the impact category k_{unique} and reduces equation (1) by all summands except one. The overall mathematical calculation in the LCA can be expressed as follows:

Linking of the emission factor to an inventory amount using the proxy:

$$EF_{k_{unique}} * proxy_{k_{unique}i_{unique}} = m_{i_{unique}} \quad (4)$$

Calculating characterization results for the impact category k_{unique}:

$$m_{i_{unique}} * LCIA_{i_{unique},k_{unique}} = E_{k_{unique}} \quad (5)$$

Prove of the emission factor EF_k being equal to the characterization results E_k for the impact category k, by inserting (4) in (5) and using (3):

$$EF_{k_{unique}} * proxy_{k_{unique}i_{unique}} * LCIA_{i_{unique},k_{unique}} = E_{k_{unique}}$$

$$EF_{k_{unique}} * \frac{1}{LCIA_{i_{unique},k_{unique}}} = \frac{E_{k_{unique}}}{LCIA_{i_{unique},k_{unique}}}$$

This is how the emission factor in the LCI can be displayed as a characterization result after executing the LCIA.

The EF 3.0 which is required by EN 15804 for EPDs comprises 19 impact categories, 15 of which have a unique corresponding ecosphere flow i_{unique}. When no unique ecosphere flows are available, two workable solutions exist to address this issue: either a balancing mechanism (chapter 3.2) or a summation of other impact category results (chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**), dependent on the impact category's structure, may be employed.

3.2. Balancing Mechanism for Non-Unique Ecosphere Flows

In the absence of any unique ecosphere flow within the impact category $k_{non-unique}$, the utilization of proxies as previously outlined would increase another impact category $k_{to\ be\ balanced}$. The inventory result m_i consists of two parts, the results for impact category $k_{to\ be\ balanced}$ with its unique ecosphere flow i_{unique} and the part from the non-unique ecosphere flow $i_{non-unique}$ of the impact category $k_{non-unique}$. This can be expressed as follows for the impact category $k_{to\ be\ balanced}$:

$$\sum_k EF_k * proxy_{k,i} = m_i \quad (6)$$

$$EF_{k_{to\ be\ balanced}} * proxy_{i_{unique},k_{to\ be\ balanced}} + EF_{k_{to\ be\ balanced}}$$

$$* proxy_{i_{non-unique},k_{to\ be\ balanced}} = m_{i_{unique}} + m_{i_{non-unique}}$$

The performance of the LCIA using (1) will return:

---

[2] https://github.com/Vanessa-IREES/Finding-unique-biosphere-flows.git

$$E_{k_{to\ be\ balanced}} = LCIA_{i_{unique}k_{to\ be\ balanced}} * m_{i_{unique}} + LCIA_{i_{non-unique}k_{to\ be\ balanced}}$$

$$* m_{i_{non-unique}}$$

Or expressed in a simpler formula:

$$E_k = correct\ characterization\ result\ +\ increase$$

In order to achieve balance, the unintentional increase must be offset by a negative amount to reach a net of zero. It can be expressed as follows:

$$E_k = correct\ characterization\ result\ +\ increase - balancing =$$
$$correct\ characterization\ result \quad (7)$$

An example can be observed concerning the impact categories of acidification and photochemical ozone formation. In the context of acidification ($k_{non-unique}$), it is impossible to identify a unique ecosphere flow. However, the ecosphere flow "sulphur dioxide" (SO₂), which is also included in the impact category photochemical ozone formation $k_{to\ be\ balanced}$, can be considered as representative ecosphere flow $i_{non-unique}$ in the $proxy_{acidification,SO_2}$. By utilising SO₂ in the proxy, it is possible to obtain a correct characterization result for the impact category acidification $E_{acidification}$. Furthermore, SO₂ is recognised by the impact category of photochemical ozone formation $k_{to\ be\ balanced}$, which will there lead to an increase in the results. The impact category photochemical ozone formation already has a unique ecosphere flow i_{unique}, the non-methane volatile organic compounds from unspecified origin (NMVOC) and is already correctly represented by t$proxy_{i_{unique},k_{to\ be\ balanced}}$. The utilisation of the proxy for acidification with the non-unique ecosphere flow SO₂ for acidification has the consequence of producing incorrect results in photochemical ozone formation.

This is why the proxy of the impact category $k_{non-unique}$, in the example of acidification, requires a balancing in the impact category $k_{to\ be\ balanced}$, in the example of photochemical ozone formation. In order to achieve a balancing of the impact category $k_{to\ be\ balanced}$, it is necessary to introduce a second unique flow $i_{other\ unique}$ in the impact category $k_{to\ be\ balanced}$, with a $proxy_{k_{to\ be\ balanced},i_{other\ unique}}$. To return to the previous example, a secondary unique ecosphere flow is required in the photochemical ozone formation, in order to balance the increase coming from SO₂ of acidification. Consequently, a negative amount, representing a fictious environmental uptake, of a substance such as butanol is necessary to reduce the results of the photochemical ozone formation. As butanol is a unique ecosphere flow of photochemical ozone formation, no other impact categories are affected. This is illustrated in Figure 2.
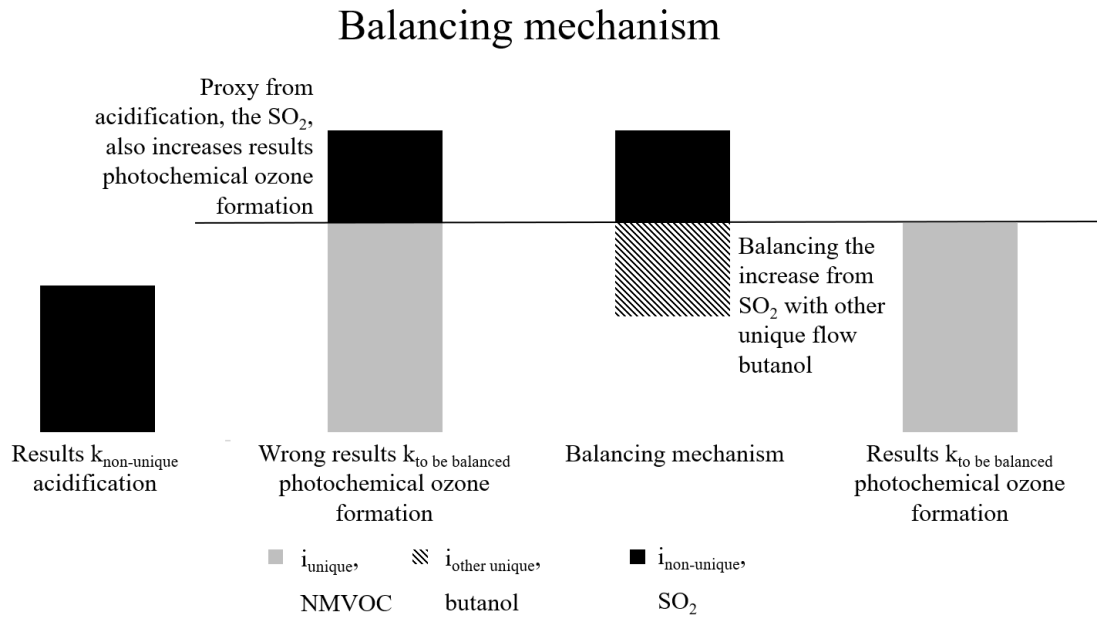
Figure 2. Graphical explanation of the balancing mechanism of impact assessment results which is used for impact categories without unique ecosphere flows, in the example of acidification which is why the photochemical ozone formation is balanced.

The $proxy_{k_{non-unique},i_{non-unique}}$ for the balancing mechanism is derived using the following formulas (1), (3), (5) and (7):

$$increase + balancing = 0$$

$$balancing = m_{i_{other\ unique}} * LCIA_{i_{other\ unique},k_{to\ be\ balanced}}$$

$$increase = m_{i_{non-unique}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}}$$

$$m_{i_{non-unique}} = EF_{k_{to\ be\ balanced}} * proxy_{k_{to\ be\ balanced},i_{non-unique}}$$

$$m_{i_{other\ unique}} = EF_{k_{to\ be\ balanced}} * proxy_{k_{to\ be\ balanced},i_{other\ unique}}$$

$$proxy_{k_{to\ be\ balanced},i_{non-unique}} = \frac{1}{LCIA_{i_{non-unique},k_{non-unique}}}$$

The proxy for the other unique ecosphere flow $proxy_{k_{to\ be\ balanced},i_{other\ unique}}$ can then be derived as follows:

$$-increase_{k_{to\ be\ balanced}} = balancing_{k_{to\ be\ balanced}}$$

$$-m_{i_{non-unique}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}} =$$

$$m_{i_{other\ unique}} LCIA_{i_{other\ unique},k_{to\ be\ balanced}} , \ -EF_{k_{to\ be\ balanced}} *$$

$$proxy_{k_{to\ be\ balanced},i_{non-unique}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}} = EF_{k_{to\ be\ balanced}} *$$

$$proxy_{k_{to\ be\ balanced},i_{other\ unique}} * LCIA_{i_{other\ unique},k_{to\ be\ balanced}} ,$$

$$-proxy_{k_{to\ be\ balanced},i_{non-unique}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}}$$

$$= proxy_{k_{to\ be\ balanced},i_{other\ unique}} * LCIA_{i_{other\ unique},k_{to\ be\ balanced}},$$

$$proxy_{k_{to\ be\ balanced},i_{other\ unique}}$$

$$= \frac{-\ proxy_{k_{to\ be\ balanced},i_{non-unique}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}}}{LCIA_{i_{other\ unique},k_{to\ be\ balanced}}},$$

$$proxy_{k_{to\ be\ balanced},i_{other\ unique}}$$

$$= -\frac{\frac{1}{LCIA_{i_{non-unique},k_{non-unique}}} * LCIA_{i_{non-unique},k_{to\ be\ balanced}}}{LCIA_{i_{other\ unique},k_{to\ be\ balanced}}},$$

$$proxy_{k_{to\ be\ balanced},i_{other\ unique}} = -\frac{\frac{LCIA_{i_{non-unique},k_{to\ be\ balanced}}}{LCIA_{i_{non-unique},k_{non-unique}}}}{LCIA_{i_{other\ unique},k_{to\ be\ balanced}}} \quad (7)$$

The increase can be derived from the characterization factors of the non-unique ecosphere flows in the impact categories $k_{to\ be\ balanced}$ and $k_{non-unique}$ and from the characterization factor of the other unique ecosphere flow in the impact category $k_{to\ be\ balanced}$. Overall, the same methodology as described in section 3.1 is employed to calculate the proxy, but the balancing mechanism is required due to the non-unique ecosphere flow in the impact category $k_{non-unique}$.

When a non-unique ecosphere flow $i_{non-unique}$ has an impact on more than one other impact category, the formula can be extended by integrating other impact categories and their respective balancing.

The EF 3.0, which is used for EPDs in OBD, has three impact categories, where the balancing mechanism must be applied. The chosen ecosphere flows for these impact categories are presented in Table 2.

Table 2. Impact categories affected by the balancing mechanism in the OBD.

Impact category	$i_{non-unique}$	$i_{other\ unique}$
Acidification potential, Accumulated Exceedance (AP)	Sulphur dioxide ('air', 'lower stratosphere + upper troposphere')	Butanol ('air', 'low population density, long-term')
Eutrophication potential - terrestrial	Nitrate ('air')	Nitrate ('water, ground')
Depletion potential of the stratospheric ozone layer (ODP)	Methane, dichlorodifluoro-, CFC-12 ('air', 'low population density, long-term')	Carbon dioxide, fossil ('air', 'urban air close to ground')

3.3. Impact Category k as Sum of Other Impact Categories

If a specific characterization result of an impact category *k* is the sum of other characterization results, the proxy for the specific impact category can be set to zero. E.g., the GWP100 in EN 15804 is comprised of the GWP100, fossil, the GWP100, biogenic and the GWP100, luluc. Consequently, the GWP100 total does not need to have a proxy. The characterization result is correctly displayed by the proxies for GWP100, fossil, GWP100 biogenic and GWP100, luluc and also accounts for the GWP100.

$$E_{GWP100} = E_{GWP100\ fossil} + E_{GWP100\ biogenic} + E_{GWP100\ luluc}$$
$$proxy_{k_{sum},i} = 0$$

3.4. Technical Implementation in Python

This section outlines how to import the products and their respective emission factors from the OBD database, which is stored as a csv file, into the LCI of the LCA program Brightway. The code is available on GitHub[3]. The CSV file of the OBD database is parsed into a data frame using the Python pandas library (The pandas development team 2020). To facilitate data processing, German names are employed where the English name is missing and in certain cases non-numeric characters, which indicate different types of data gaps, are substituted with 0.

To enhance user-friendliness, the possibility of aggregation of modules and the direct visibility of scenarios have been integrated into the technical import structure. The configurable number of modules in the aggregation allows for the inclusion of a specific subset of modules, such as only module A, the construction phase. Similarly, it is also possible to include the aggregate of all modules associated with a particular product specified in a scenario. To achieve this aggregation, the import process includes a series of steps to standardize the data provided by the CSV file. This standardization facilitates subsequent calculations and the further processing of the data:

1. Labelling scenarios

UUIDs of products do not provide insight into the presence of scenarios in the data set. Aggregating solely by UUID would include all scenario-specific variations, leading to incorrect results, as distinguishing between different scenarios of a single product with one UUID would not be possible.

To resolve this, a unique composite key is built by combining a scenario with the product and its UUID. Subsequently, these UUID-scenario tuples serve as the base for further aggregation. This is achieved by the insertion of a new column which bears the concatenation of the 'UUID' column and the 'Szenario' or scenario column. Unmarked rows, i.e. rows which do not bear a specific scenario identifier, are used to fill up the scenario rows. As an example, a product may be associated with two scenarios, 1 and 2, which are represented by two rows for module: 'C', and one row for the modules 'A' and 'B' belonging to both scenarios. It is then assumed that modules 'A' and 'B' are identical for both scenarios. In this example product, the import script will copy the unmarked rows and insert them once for each scenario. This results in the scenarios effectively having rows labelled 'A,B,C'.

2. Inconsistent data entries for life cycle stage modules

The OBD data entries of life cycle stage modules vary considerably between products. Not all life cycle stage modules are available. One product might contain data for module 'A1' and 'B', another for 'A1' and 'C', and a third one for 'A1-A3'. To ensure consistency, each product-scenario-tuple must be provided with the modules that are currently absent. The newly inserted rows for each life cycle stage module have a value of 0 as the emission factor. This process standardizes the data structure across all products and modules, enabling automated processing and comparability. As a result, users can perform LCA calculations with all 471 products from the OBD, even when some products lack data for certain modules, without requiring manual adjustments. However, the insertion of 0 values does not indicate that there are no emissions in these modules. It simply serves to ensure a standardized basis for the comparison of existing data.

3. Aggregation of life cycle stage modules 'A1-A3'

The OBD includes the aggregated modules 'A1-A3' for certain products and the disaggregated modules 'A1', 'A2' and 'A3' for others. Since the 'A1-A3' module lacks weighting information, they cannot be disaggregated to 'A1', 'A2' and 'A3'. Consequently, the import script aggregates the individual modules 'A1', 'A2' and 'A3' when they are provided for a product. This results in the data

---

[3] https://github.com/m0e161/brightway-ef4lca.git

being structured in a consistent manner, which simplifies automated data processing. This yields a data frame comprising the aggregated module 'A1-A3' for all products.

The resulting data frame, containing the enhanced OBD data, can now be imported into Brightway2. Each product-scenario-tuple is imported and is required to have at least a 0 or an emission factor for each of the life cycle stage modules 'A1-A3', 'A4', 'A5', 'B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'C1', 'C2', 'C3', 'C4' and 'D'.

Further custom aggregations may be applied before importing with either the Pandas/Python default means or the provided 'custom_aggregate' function. The 'custom_aggregate' function will insert a specific key into the resulting data frame. This key contains the instructions for the aggregation, for example, the key 'A1-A5', indicates that the modules 'A1-A3', 'A4', and 'A5' should be aggregated. The aggregation function is based on a fixed module order which is defined in the configuration. The 'custom_aggregate' function is primarily used to reduce the modelling effort. The grouping of modules generates one single activity, whereas the ungrouped modules indicate one activity per module. Consequently, for grouped modules, only one activity is required to be selected during modelling, which is less time-consuming.

The goal of the LCI is to hand over the correct inventory amounts m_i to the LCIA to be able to calculate LCA results. Inventory amounts m_i from EPDs as data source can be calculated by multiplying the emission factor EF_k and the $proxy_{k,i}$ as described in formula (6). The proxies are stored in a database.

This proxy database links the emission factors for each impact category to ecosphere flows, allowing for the calculation of the LCIA. The proxy ecosphere flows have been carefully selected as described in chapter 3.1., 3.2 and 3.3. The setup of the proxy database is achieved through a manual configuration, which defines the mapping of an OBD impact category *k* to an array of ecosphere flows *i* and the corresponding amount. This corresponding amount is calculated using the three solutions for proxy calculation as described in chapters 3.1, 3.2 and 3.3. For the first case of a unique flow in an impact category, the corresponding amount is the reciprocal characterization factor as described in formula (3) in chapter 3.1. For the second case of the impact category not consisting of any unique ecosphere flows and the employment of the balancing mechanism, the corresponding amount is calculated using formula (7) from chapter 3.2. For the third case of the impact category being the sum of other impact categories, the value of zero is the corresponding amount as described in chapter 3.3. Subsequently, each corresponding amount of ecosphere flows *i* is imported as a Brightway activity, with links to the ecosphere. The ecosphere in Brightway is represented by the biosphere3 database and has been pre-implemented in Brightway. The proxy database is not only linked to ecosphere flows but also to impact categories in the OBD. Therefore, the data frame which contains the standardised and potentially aggregated OBD data, is imported in such a way that each product, scenario, module (and aggregation) combination is imported as a Brightway activity. These OBD Brightway activities have exchanges to the proxy database activities for each impact category.

3.5. Example of LCA Calculations with Data from Process-Based and EPD Databases

The technical building equipment of a tiny wooden house is modelled to provide an example for using data from process-based and emission factor databases. The house is intended to be used as a learning room for students at a university in Germany. Therefore, a complete LCA of the technical building equipment is conducted to be able to compare technical configurations. This 9 m² room is fully independent of the sewer system as it has no bathroom. At the end, one component coming from OBD database is analysed in detail to identify the precision of the method.

The functional unit is the technical functionality of a building including a warm space, electricity, light and smoke detection per year. The modules considered are module A, the production of the materials for the equipment without the transport to the construction site, and module C, the end-of-life treatment of the technical building equipment for the technical building equipment. Additionally, the heating demand is considered as part of module B, the use phase, and is assumed to be 3900 kWh/year. The user's electricity is excluded as it is highly uncertain how the users in the

study room behave. OBD data is used as a primary data source, as it is the standard LCA database for the construction sector in Germany. To close data gaps, the ecoinvent database 3.9 cutoff is used. In the OBD database, six out of 13 parts of the technical building equipment can be found. For six components, ecoinvent datasets and for the heating system a combination of OBD and ecoinvent datasets are used.

The components, which are unavailable in the OBD database are a communication module, a control cabinet, a display (5"), a motion sensor and a smoke detector. These components are modelled using data from ecoinvent. Furthermore, heat pumps in the OBD do not indicate which electricity mix is used to operate the heat pump. Neither the background system is visible, as the OBD is an emission factor database, nor the documentation indicate the used electricity mix. As it is expected that the electricity consumption for the heat pump has a large influence on the results, ecoinvent data is used to model the heating system. The background system of heat pumps in ecoinvent is transparent and the exact electricity demand and the electricity origin are visible. Therefore, the German electricity mix from 2019, which is available in ecoinvent, is included to run the heat pump in the model. A self-sufficiency ratio to run the heat pump with electricity from PV and the electricity storage of 25% is assumed after (Bee et al. 2019).

The lifetime of the components is taken from the European Commission and Joint Research Centre (2021) and for the PV panel form Fraunhofer ISE (2024). The lifetimes as well as the components and their data source are displayed in Table 3.

Table 3. Components of the technical building equipment of the tiny house and their data sources in the databases.

Technical building equipment	Database used	Amount	Name of the dataset	Lifetime
Cable duct	OBD	34.56 kilogram	Cable duct PVC, rigid; PVC	30 years
Communication module	eccoinvent	1 unit	Internet access equipment	15 years
Control cabinet	ecoinvent	2 kilograms	electronics, for control units	15 years
Charge controller	ecoinvent	1 unit	battery management system, for Li-ion battery	15 years
Display (5")	ecoinvent	0.29 units	display, liquid crystal, 17 inches	15 years
Heating system	ecoinvent	14040 MJ	heat production, air-water heat pump 10kW with 75% (1087 kWh) electricity from the German grid (market for electricity, low voltage, DE, ecoinvent) and 25% (326 kWh) from PV and the storage	per year

	OBD	9 m²	Underfloor heating system PEX (200mm distance); 200 mm distance	20 years
Motion sensor	ecoinvent	0.27 kilograms	polycarbonate	15 years
		0.03 kilograms	electronic component, passive, unspecified	15 years
Smoke detector	ecoinvent	3*0.15 kilograms	polystyrene, extruded	15 years
		3*0.05 kilograms	electronic component, passive, unspecified	15 years
Power outlet	OBD	7 units	Electric socket; 1 piece	30 years
Lightning	OBD	2 units	Rocker lightswitch; 1 piece	30 years
		5 units	Louvrelight 2x T8-36W (LFL); 1 piece	15 years
		1 unit	LED office luminaire	15 years
Inverter	OBD	1 unit	inverter production, 2.5kW	15 years
PV	OBD	1.98 m²	Photovoltaic system 1000 kWh/m²*a	20 years
Electricity storage	OBD	12 kWh	Lithium iron phosphate (LiFePO4) battery (per 1kWh storage); 1kWh storage capacity	15 years

The results are calculated using the impact assessment method EF 3.0 EN15804. In Figure 3, the relative results per impact category and component are visible. Results displayed in blue represent data from OBD[4], results in red data from ecoinvent and results in green mixed data from OBD and ecoinvent. For all components from OBD besides the inverter, no data is available for the impact categories ecotoxicity freshwater, human toxicity carcinogenic and non-carcinogenic, ionising radiation: human health and land use. The results of these impact categories are incomplete, which is why they are not displayed in the following figures.

---

[4] Please note, that the OBD version from 2023 has been used.
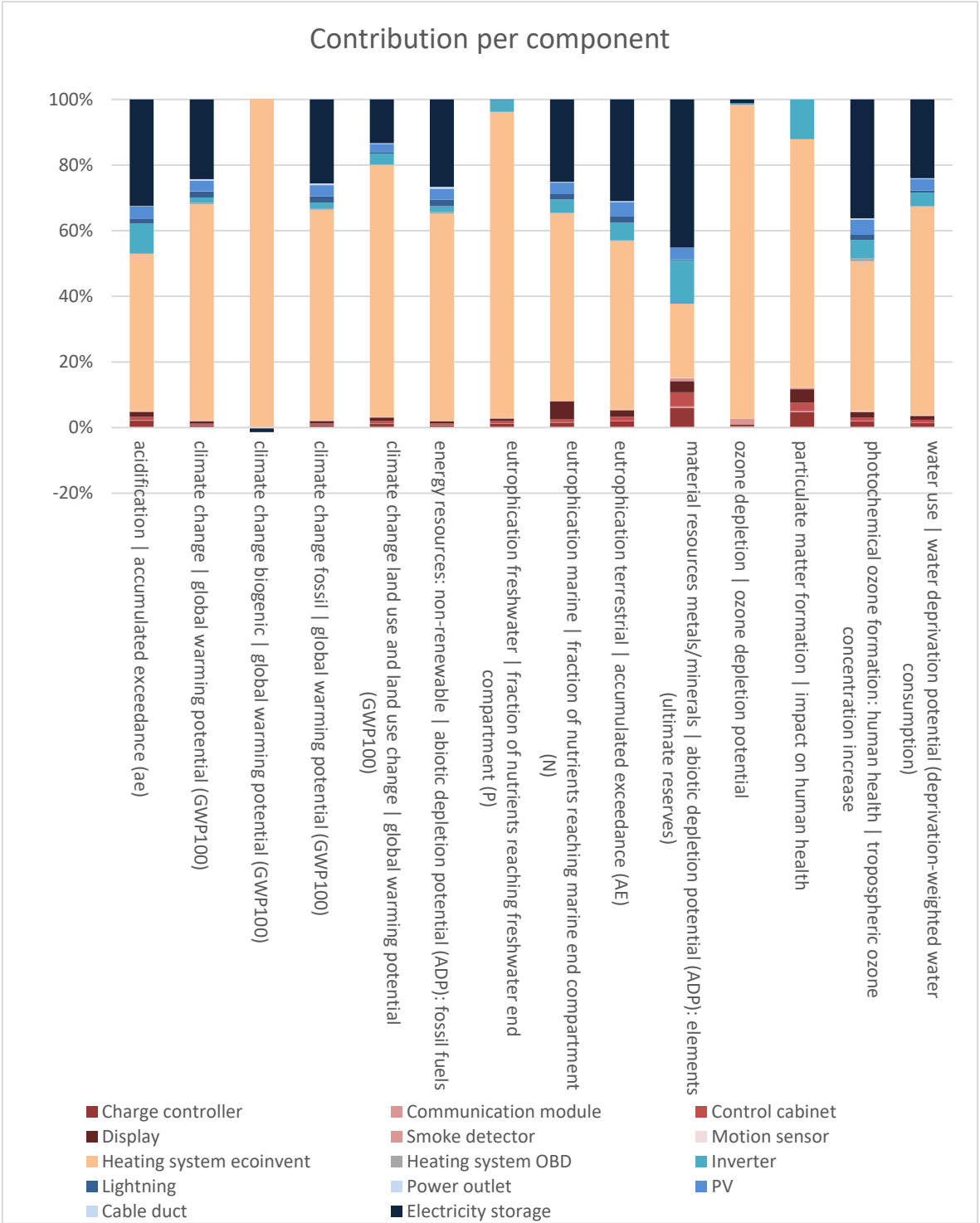
Figure 3. Results per impact category and component, in blue data from OBD and in red data from ecoinvent.

In general, the heating system contributes the most to all impact categories. It consists of data from the OBD and the ecoinvent database. When analysing the GWP100 total of the heating system in more detail, it can be recognized that the underfloor heating system with data from OBD contributes to 0,5% of the GWP100 total, whereas the heat from the heat pump with ecoinvent data contributes to 65.3% of the GWP100 total. Almost all of the contributions to the GWP100 from the heating system are derived from ecoinvent data. This is similar to other impact categories and increases the share of the contribution from ecoinvent data. The contribution of each database to the results per impact category is displayed in Table 4.

Table 4. : Contribution of ecoinvent and OBD data to the results.

Contribution to the results per database	OBD		ecoinvent	
acidification | accumulated exceedance (ae)	47.08%	1.57 mol H+-Eq	52.92%	1.77 mol H+-Eq
climate change | global warming potential (GWP100)	31.96%	310.51 kg CO2-Eq	68.04%	661.13 kg CO2-Eq
climate change biogenic | global warming potential (GWP100)	-0.99%	-0.47 kg CO2-Eq	100.99%	47.48 kg CO2-Eq
climate change fossil | global warming potential (GWP100)	33.61%	310.43 kg CO2-Eq	66.39%	613.16 kg CO2-Eq
climate change land use and land use change | global warming potential	19.87%	0.21 kg CO2-Eq	80.13%	0.83 kg CO2-Eq
energy resources: non-renewable | abiotic depletion potential (ADP): fossil	34.81%	4074.45 MJ, net	65.19%	7629.18 MJ, net
eutrophication freshwater | fraction of nutrients reaching freshwater end	3.78%	0.03 kg P-Eq	96.22%	0.72 kg P-Eq
eutrophication marine | fraction of nutrients reaching marine end	34.74%	0.23 kg N-Eq	65.83%	0.72 kg P-Eq
eutrophication terrestrial | accumulated exceedance (AE)	43.17%	2.46 mol N-Eq	56.83%	3.24 mol N-Eq
material resources metals/minerals | abiotic depletion potential (ADP):	62.27%	0.02 kg Sb-Eq	37.73%	0.01 kg Sb-Eq
ozone depletion | ozone depletion potential	1.62%	1.48E-06 kg CFC-	98.38%	0 kg CFC-11-Eq
particulate matter formation | impact on human health	12.13%	1.25E-06 disease	87.87%	9.02E-06 disease
photochemical ozone formation: human health | tropospheric ozone	49.32%	0.91 kg NMVOC-	50.68%	0.94 kg NMVOC-Eq
water use | water deprivation potential (deprivation-weighted water	32.69%	58.52 m3 world	67.31%	120.5 m3 world eq.

The heating system contributes the most to all impact categories, which is why its data quality is analysed in more detail. It must be recognized that the seasonal performance factor of the heat pump in ecoinvent, which is 2.8, is from 1998, which will result in an overestimation of the share of

the heating system. Additionally, the data quality of the heat pump is influenced not only by the temporal accuracy, but also by assumptions for the electricity demand. For the heat pump, the assumed self-sufficiency ratio of 25% is likely to be higher due to the large electricity storage. This would decrease the impact resulting from the grid electricity consumption of the heat pump. Due to the highly uncertain electricity consumption of the users, which is excluded in this example, the electricity produced by the PV plant could be needed by the users and therefore is unavailable any more for operating the heat pump. Therefore, a self-sufficiency ratio of 25% of the heat pump is assumed to be sufficiently valid. The data quality of assumptions from the electricity demand and the temporal accuracy of the heating system datasets from OBD and ecoinvent are sufficiently illustrating this example of process-based and emission factor databases in one LCA calculation.

To exemplarily analyse the precision of the method, one component is looked at in detail. The original ecoinvent dataset of the Lithium iron phosphate (LiFePO4) battery (per 1kWh storage); 1kWh storage capacity is compared to the results in the Activity-Browser, because it contributes significantly to the results of the example. The results in the Activity-Browser are generated by using the method described above and deviate from the original ones in the range of 10^{-5} to 10^{-9} depending on the impact category. The deviation per impact category is displayed in Table 5 with the impact categories underlined where a balancing mechanism is performed. The highest deviation in percent from the absolute values is 0.8% for the ozone depletion followed by 0.003% for eutrophication: freshwater. The lowest deviation from the absolute results is $9.7 \cdot 10^{-12}$ for the energy ressources: non-renewable, whereas all other results deviate in the range of 10^{-8} and 10^{-9}.

Table 5. Deviation of results in absolute values displayed in the Activity-Browser from the original OBD datasets with the impact categories underlined where a balancing mechanism is performed.[5].

Deviation of results displayed in the Activity-Browser from the original OBD datasets of the product Lithium iron phosphate (LiFePO4) battery (per 1kWh storage); 1kWh storage capacity	
<u>acidification | accumulated exceedance (ae)</u>	<u>7.2 E-08</u>
climate change | global warming potential (GWP100)	2.7 E-05
climate change biogenic | global warming potential (GWP100)	3.4 E-09
climate change fossil | global warming potential (GWP100)	2.7 E-05
climate change land use and land use change | global warming potential (GWP100)	3.9 E-09
energy resources: non-renewable | abiotic depletion potential (ADP): fossil fuels	3.8 E-08
eutrophication freshwater | fraction of nutrients reaching freshwater end compartment (P)	2.7 E-08
eutrophication marine | fraction of nutrients reaching marine end compartment (N)	5.1 E-09
<u>eutrophication terrestrial | accumulated exceedance (AE)</u>	<u>6.1 E-09</u>
material resources metals/minerals | abiotic depletion potential (ADP): elements (ultimate reserves)	1.5 E-08
<u>ozone depletion | ozone depletion potential</u>	<u>1.1 E-08</u>
photochemical ozone formation: human health | tropospheric ozone concentration increase	3.1 E-08
water use | water deprivation potential (deprivation-weighted water consumption)	5.4 E-08

The deviation of the results being in the range of 10^{-8} for the impact categories with a balancing mechanism might indicate, that performing a balancing mechanism is unaffecting the precision of

---

[5] Please note, that the OBD version from 2023 has been used.

the displayed results. The GWP100 total as well as the GWP100 fossil show the lowest precision of all impact categories. The GWP100 total is calculated by summing up GWP100 biogenic, GWP100 fossil and GWP100 land use and land use change, which is why the lowest precision of all summands determines the precision of the GWP100 total. After analysing the code importing the data in Brightway, the data is given with a much higher precision to Brightway than shown in Table 5. Additionally, Brightway does not round when importing data. The only case, where Brightway is rounding to at least one decimal point, is the export of data to the Ecospold1 XML format (Brightway Developers 2024). This means that the pandas library is decreasing the precision when handling .csv data formats, which has been reported in the Stackoverflow forum (beta 2017). In general, these differences in LCA calculations are very small. The results displayed in the Activity-Browser represent the original OBD datasets with sufficient precision.

This example shows the simple use of a conventional process-based database, the ecoinvent database, together with an emission factor database, the OBD, in one LCA calculation. The presented methodology is used to import the OBD database in the LCA program, the Activity-Browser, to calculate an LCA using data from OBD and ecoinvent simultaneously. It shows that the methodology allows the filling of data gaps by using process-based databases and emission factor databases at the same time.

4. Discussion: Limitations of the Method and the Example

The use of the presented method allows data gaps in LCAs to be filled with emission factors. While it is reasonable to avoid combining data from different databases, given that assumptions and data modelling rules are typically harmonised within a single database but not across different databases, it is nevertheless rational to employ additional databases to fill data gaps when one database is insufficient. The combined utilisation of these databases results in a more extensive dataset.

A drawback of the proposed method of integrating emission factors into LCA is that it allows the utilisation of any impact assessment method within the LCIA. Emission factors are LCA results that are clearly linked to one particular impact category. When another impact category is used when making available emission factors in LCA programs employing the proposed method, it results in issues. The integration of these emission factors using alternative impact categories is simply incorrect. The consistency between impact categories of providing emission factors and integrating them as data source in LCA has already been discussed by Strazza et al. (2016). The presented method employs the use of a proxy to assign a substitute ecosphere flow to every emission factor in the LCI. This ecosphere flow is a fictional flow; rather it is a fictitious construct that serves as a workaround to make data available. This fictitious ecosphere flow enables the application of any available impact category for result calculation, but only one is correct. The impact category used for the data provision of emission factors must be the same as the one used for the data integration in LCA. As an example, our method allows the calculation of OBD results using GWP100 after IPCC 2021, which is an incorrect approach given that the OBD is only valid for GWP100 total after EF 3.0, EN15804. Users must be aware that this method is only applicable for corresponding impact categories. User awareness avoids application errors and ensures the great benefit through providing usable emission factor data.

Upon displaying the results of the OBD in the Activity-Browser, the contribution analysis does not display any contributions for the OBD data. This is due to the absence of process information in emission factor databases such as the OBD and will happen for any other emission factor database as well. In order to make such emission factor databases without process information available in the LCI, proxies are used to assign a representative ecosystem flow to each emission factor in the LCI. This representative ecosphere flow functions solely as placeholder without meaningful or interpretable content. Furthermore, a contribution analysis for multiple products from emission factor databases within a single LCA model is feasible. A contribution analysis would at least indicate the share of impacts attributable to a single emission factor, even though process-based information

is not available. A Monte Carlo Simulation to assess the uncertainty of the LCA is infeasible, because emission factor databases lack uncertainty information. Even though the analysis of results is limited when using emission factors, the benefit of fewer data gaps outweighs this disadvantage.

The most significant limitation associated with the utilisation of emission factors in LCA is missing data transparency. The lack of transparency in the process structure prevents an assessment of the quality of the data. Consequently, the quality of the emission factors is unclear and modelling errors and inaccurancies are invisible. This is why, only validated LCA data, such as EPD data, should be imported.

Furthermore, the utilisation of validated EPD data in LCA has weaknesses, which can be attributed to the EPD standard EN 15804. Firstly, the scope of the data is frequently unclear. The impact categories that must be used when publishing an EPD in accordance with EN 15804 must originate from the European Commission's EF methodology. However, the version of the EF is not indicated. It is therefore necessary to request the EF version of the EPD before integrating the data into the LCA. Secondly, not all environmental impact categories are mandatory in the EPD. It needs to be considered, that in some OBD datasets impcact categories are missing. These should be left out in the results, as demonstrated in the example in chapter 3.5. This results in a lack of information in a maximum of 7 out of 19 impact categories. It must be evaluated whether the more specific EPD data outweighs the lack of information in the additional impact categories.

One issue of utilising our method for the OBD database is observed when the OBD data is processed. The OBD database provides not only LCIA results but also end-of-life scenarios, which make the import progress more challenging. Unfortunately, these scenarios cannot be immediately identified, since the names of the scenarios names vary considerably. Additionally, numerous ways of indicating data gaps or data outside the system boundaries exist, such as NaN or 0. For OBD users this supplementary information might be crucial, but increases the effort of digital data management. Data preprocessing is inevitable due to end-of-life scenarios and the form of storing data gaps. Nevertheless, this issue comes from the data quality of the raw data and does not hinder in applying the methodology.

Another deficit is the lack of any process-based background information concerning efficiency or hazardous materials within emission factors. This lacks actuality and accuracy of the background information. This is a general issue when utilising EPD as a data source as Strazza et al. (2016) also mention, and cannot be solved with this methodology. As Mendoza Beltran et al. (2020) demonstrate, background information can be highly relevant, particularly in the context of changes in the electricity mix within an electricity-intensive production process. To counteract this, it is essential that emission factors are updated regularly and that the most up-to-date emission factor data is employed. The utilisation of emission factor data instead of generic process-based data can even be more accurate when current and manufacturer-specific data, for example as EPD, is utilised.

5. Conclusions and Recommendations: Advantages of the Method and Outlook

The integration of the OBD database using proxies allows the integration of OBD products in LCA with products from other process-based databases such as the ecoinvent database. This increases user-friendliness as only one LCA modelling and calculation is required when using emission factors without process-based information and databases with process-based information. In addition, a lot of other data, e.g. from EPD databases such as OBD, are easily available for LCA calculations.

To facilitate the utilisation of OBD, it would be beneficial to improve certain aspects of the database. Enhancing the user-friendliness of the OBD could be achieved by providing standardised textual descriptions of end-of-life scenarios and EPD system boundaries. This would reduce the effort required for data processing and would ensure the provision of transparent information.

In contrast to earlier methodologies, the presented methodology employs the LCA program structure as illustrated in Figure 1, which facilitates the processing of LCIA results. Earlier methods require post-processing of the results, as the program displays two separate results. The method

under discussion here implements the emission factor databases in an earlier LCA calculation step, the LCI, and leaves the LCIA with its impact categories untouched. This allows the results of emission factor databases and conventional process-based databases to be summed up by the program, thus producing a single result for each LCA calculation.

Furthermore, this approach permits an uncomplicated evaluation of the results within the software. When conducting an LCA, comparing two absolute values per impact category for two products is informative, but so too is the hotspot analysis, which identifies the most influential parts of a product in the impact assessment. The performance of hotspot analyses for two results in one LCA is inconvenient, which is the case for existing methods of emission factor integration in LCA. The performance of analyses and the user-friendliness of the process are enhanced when only one result per LCA calculation is obtained. However, the analysis functions can be limited when analysing results from emission factor databases due to the lack of background process data.

Additionally, our approach offers the benefit of providing access to more specific data. In addition to regional data, the LCA can also integrate supplier-specific data from and for environmental reporting. As illustrated in the example, the OBD provides more specific and regional data for the calculation of environmental impacts. Such data can be utilised with minimal effort by employing the proposed methodology and integrating it into the Brightway and Activity-Browser structure. The available data can help to fill data gaps and to use more suitable data, which increases the quality of the results.

By the general validity of this approach the method is not restricted to a specific set of data sources, emission factors or LCA programs. It permits the general integration of LCA emission factors in the form of impact category results without process-based background data into LCA. The approach is not limited to the previously presented example of the OBD integration; however, this example provides a foundation for integrating additional emission factors into general LCA software. Accordingly, the CSV file of the OBD, which is read by the Python code to integrate it into Brightway, could serve as a template for the inclusion of further EPD data. The code provided allows for the integration of all environmental declarations with results for the impact categories as specified in EN 15804. Furthermore, this approach is universally applicable. Users may employ this method to transform any database into a format readable by any LCA program allowing the integration of emission factors into the software. The method is not limited to a specific example and allows for the integration of any emission factor data without process-based background information in LCA programs. The open-source code will contribute to an increase in the number of users and support other milestones, enhancing data availability in LCA.

Data availability statement

The code supporting the findings of this study is openly available in two GitHub repositories:

- Finding unique biosphere flows of an impact assessment method: https://github.com/Vanessa-IREES/Finding-unique-biosphere-flows.git

- Integrating emission factors in LCI databases – brightway-ef4lca: https://github.com/Vanessa-IREES/brightway-ef4lca.git

References

Abu-Ghaida H, Ritzen M, Hollberg A, Theissen S, Attia S, Lizin S (2024) Accounting for product recovery potential in building life cycle assessments: a disassembly network-based approach. Int J Life Cycle Assess 29:1151–1176. https://doi.org/10.1007/s11367-024-02324-8

Baehr J, Zenglein F, Sonnemann G, Lederer M, Schebek L (2024) Back in the Driver's Seat: How New EU Greenhouse-Gas Reporting Schemes Challenge Corporate Accounting. Sustainability 16:3693. https://doi.org/10.3390/su16093693

Bee E, Prada A, Baggio P, Psimopoulos E (2019) Air-source heat pump and photovoltaic systems for residential heating and cooling: Potential of self-consumption in different European climates. Build Simul 12:453–463. https://doi.org/10.1007/s12273-018-0501-5

beta (2017) Pandas read csv file with float values results in weird rounding and decimal digits. Stack Overflow

Brightway Developers (2024) bw2io.export.ecospold1 — Brightway documentation. https://docs.brightway.dev/en/latest/content/api/bw2io/export/ecospold1/index.html#module-contents. Accessed 20 Dec 2024

Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen (2024) ÖKOBAUDAT. https://www.oekobaudat.de/. Accessed 22 Feb 2024

de Villiers C, La Torre M, Molinari M (2022) The Global Reporting Initiative's (GRI) past, present and future: critical reflections and a research agenda on sustainability reporting (standard-setting). Pacific Accounting Review 34:728–747. https://doi.org/10.1108/PAR-02-2022-0034

Del Rosario P, Palumbo E, Traverso M (2021) Environmental Product Declarations as Data Source for the Environmental Assessment of Buildings in the Context of Level(s) and DGNB: How Feasible Is Their Adoption? Sustainability 13:6143. https://doi.org/10.3390/su13116143

Deutsches Institut für Normung (2022) DIN EN 15804:2022-03, Nachhaltigkeit von Bauwerken_- Umweltproduktdeklarationen_- Grundregeln für die Produktkategorie Bauprodukte; Deutsche Fassung EN_15804:2012+A2:2019_+ AC:2021

ecoinvent (2023) Releases Overview. https://support.ecoinvent.org/releases-overview. Accessed 28 Oct 2024

Emara Y, Ciroth A (2014) The database ÖKOBAU.DAT in openLCA and SimaPro

European Commission, Joint Research Centre (2021) Level(s) indicator 2.1: Bill of Quantities, materials and lifespans. https://susproc.jrc.ec.europa.eu/product-bureau/sites/default/files/2021-01/UM3_Indicator_2.1_v1.1_34pp.pdf. Accessed 17 Oct 2024

Figl H, Kusche O (2023) ÖKOBAUDAT-Handbuch Technisch/formale Informationen und Regeln zur ÖKOBAUDAT-Datenbank Version 2.1. https://www.oekobaudat.de/fileadmin/downloads/2023-11-20_OEBD-Handbuch_v2.1_Red_2023-12-18.pdf. Accessed 8 Jul 2024

Fraunhofer ISE (2024) Photovoltaics Report. In: Fraunhofer-Institut für Solare Energiesysteme ISE. https://www.ise.fraunhofer.de/de/veroeffentlichungen/studien/photovoltaics-report.html. Accessed 28 Oct 2024

Fritz MMC, Schöggl J-P, Baumgartner RJ (2017) Selected sustainability aspects for supply chain data exchange: Towards a supply chain-wide sustainability assessment. Journal of Cleaner Production 141:587–607. https://doi.org/10.1016/j.jclepro.2016.09.080

Greenhouse Gas Protocol (2016) Global Warming Potential Values. https://ghgprotocol.org/sites/default/files/ghgp/Global-Warming-Potential-Values%20%28Feb%2016%202016%29_1.pdf. Accessed 12 Apr 2024

Hauschild MZ, Rosenbaum RK, Olsen SI (2017) Life Cycle Assessment: Theory and Practice. Springer

Heijungs R (2022) The revised mathematics of life cycle sustainability assessment. Journal of Cleaner Production 350:131380. https://doi.org/10.1016/j.jclepro.2022.131380

Institut Bauen und Umwelt e.V. (2024) Veröffentlichte EPDs. In: IBU - Institut Bauen und Umwelt e.V. https://ibu-epd.com/veroeffentlichte-epds/. Accessed 10 Jun 2024

Intergovernmental Panel on Climate Change (IPCC) (2021) IPCC Emissions Factor Database | GHG Protocol. https://ghgprotocol.org/Third-Party-Databases/IPCC-Emissions-Factor-Database. Accessed 10 Jun 2024

International EPD® System EPD International. https://www.environdec.com/home. Accessed 10 Jun 2024

International Standardization Organization (2006a) ISO 14025:2006 - Environmental labels and declarations

International Standardization Organization (2006b) ISO 14040:2006

Meinrenken CJ, Chen D, Esparza RA, Iyer V, Paridis SP, Prasad A, Whillas E (2020) Carbon emissions embodied in product value chains and the role of Life Cycle Assessment in curbing them. Sci Rep 10:6184. https://doi.org/10.1038/s41598-020-62030-x

Mendoza Beltran A, Cox B, Mutel C, van Vuuren DP, Font Vivanco D, Deetman S, Edelenbosch OY, Guinée J, Tukker A (2020) When the Background Matters: Using Scenarios from Integrated Assessment Models in Prospective Life Cycle Assessment. Journal of Industrial Ecology 24:64–79. https://doi.org/10.1111/jiec.12825

Mutel C (2017) Brightway: An open source framework for Life Cycle Assessment. Journal of Open Source Software 2:236. https://doi.org/10.21105/joss.00236

Science Based Targets initiative (SBTi) (2024) SBTi CORPORATE NET-ZERO STANDARD

Steen B, Carlson R, Löfgren G (1995) SPINE. A Relation Database Structure for Life Cycle Assessments

Steubing B, de Koning D, Haas A, Mutel CL (2020) The Activity Browser — An open source LCA software building on top of the brightway framework. Software Impacts 3:100012. https://doi.org/10.1016/j.simpa.2019.100012

Strazza C, Del Borghi A, Magrassi F, Gallo M (2016) Using environmental product declaration as source of data for life cycle assessment: a case study. Journal of Cleaner Production 112:333–342. https://doi.org/10.1016/j.jclepro.2015.07.058

The pandas development team (2020) Pandas

Umweltbundesamt (2022) CO2-Emissionsfaktoren für fossile Brennstoffe: Aktualisierung 2022. Dessau-Roßlau

World Resources Institute and World Business Council, for Sustainable Development, GHG Protocol Initiative Team (2004) The Greenhouse Gas Protocol - A Corporate Accounting and Reporting Standard

Zargar S, Yao Y, Tu Q (2022) A review of inventory modeling methods for missing data in life cycle assessment. Journal of Industrial Ecology 26:1676–1689. https://doi.org/10.1111/jiec.13305

Zhang X, Heeren N, Bauer C, Burgherr P, McKenna R, Habert G (2024) The impacts of future sectoral change on the greenhouse gas emissions of construction materials for Swiss residential buildings. Energy and Buildings 303:113824. https://doi.org/10.1016/j.enbuild.2023.113824

(2022) Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting (Text with EEA relevance)

(2023) ISO 14068-1:2023(en), Climate change management — Transition to net zero — Part 1: Carbon neutrality

(2021) Commission Recommendation (EU) 2021/2279 of 15 December 2021 on the use of the Environmental Footprint methods to measure and communicate the life cycle environmental performance of products and organisations