LCI METHODOLOGY AND DATABASES

Making the use of scenarios in LCA easier: the superstructure approach

Bernhard Steubing[1] · Daniel de Koning[1]

Abstract

Purpose Much progress has recently been made in modelling future background systems for LCA by including future scenario data, e.g. from Integrated Assessment Models (IAMs), into life cycle inventory (LCI) databases. A key problem is, however, that this yields potentially dozens of scenario LCI databases (i.e. LCI databases that represent different scenarios and reference years), instead of a single background database, which is very impractical for LCA modelling purposes. This paper proposes an approach to overcome this problem.

Methods The approach consists of transforming all scenario LCI databases into a single superstructure database and an associated scenario difference file. The superstructure database is also a regular LCI database, but is constructed to contain all unique exchanges (elementary and intermediate flows) and processes that exist across all scenario LCI databases. The scenario difference file stores the differences between all scenarios and can be used to turn the superstructure into a specific scenario LCI database. This is very fast as it can be done in memory during LCA calculations.

Results and discussion A key advantage of the superstructure approach is that a single LCI database can be used to represent different background systems. Therefore, the practitioner does not need to re-link a foreground system to multiple LCI databases, which is work-intensive and invites modelling errors. LCA results for all scenarios and reference years can be calculated automatically. We also illustrate how the superstructure approach has been implemented in the Activity Browser open source LCA software. Although this paper introduces the superstructure approach for background scenarios, it can also be used to model foreground scenarios, and even, as implemented in the Activity Browser, combinations of background and foreground scenarios. Finally, we briefly discuss further challenges that need to be addressed for a more widespread use of background scenarios in LCA.

Conclusions The superstructure approach presents a practical solution for making the use of future background scenarios more wide-spread and, therefore, to overcome the problem of performing prospective LCA with temporally inconsistent foreground and background systems. The implementation in the Activity Browser makes the approach available for anyone and may serve as inspiration for other LCA software to implement the superstructure approach or a similar concept. While this may be an important technical milestone, additional coordination between data providers, scenario generators, LCA practitioners, and software developers will be required to further facilitate the use of background scenarios in prospective LCA studies.

Keywords Future background systems · Prospective life cycle assessment · Ex-ante LCA · LCI databases · Activity browser · LCA software · Integrated assessment models (IAM) · Shared socioeconomic pathways (SSP)

1 Introduction

Cleaner technology is required to meet future climate (IPCC 2014) and other environmental targets (Rockström et al. 2009). Many technologies are currently under development

(IEA 2020), and it will be crucial to guide the development of these technologies in order to maximize their sustainability. Life cycle assessment (LCA) is a widely used method to assess the environmental performance of technologies and their products and services (Hellweg and Canals 2014; ISO 2006). Much attention has recently been given to prospective LCA, which aims at assessing the environmental performance of technologies at a future point in time (Arvidsson et al. 2018; Cucurachi et al. 2018; van der Giesen et al. 2020; Villares et al. 2017).

✉ Bernhard Steubing
      b.r.p.steubing@cml.leidenuniv.nl

[1]   Institute of Environmental Sciences (CML), Leiden
      University, Leiden, the Netherlands

When assessing technologies at a future point in time, developments in both the foreground system (the specific technology under study) and the background system (the wider economic and technological context) should be considered in order to avoid a temporal mismatch between the specific technology and the larger technological context (Arvidsson et al. 2018; Mendoza Beltran et al. 2018; Thonemann et al. 2020; van der Giesen et al. 2020). A number of studies have shown that developments in the background system (e.g. the energy transition) may matter considerably for the outcome of LCA results (see, e.g. Cox et al. 2020; Gibon et al. 2015; Hertwich et al. 2015; Mendoza Beltran et al. 2018). Therefore, not considering such developments may result in sub-optimal recommendations to technology developers and policy makers, and may ultimately hamper progress towards environmental goals.

Scenarios for the foreground system reflect potential developments of specific technologies and can be developed by LCA practitioners in collaboration with technology developers (Tsoy et al. 2020). Scenarios for the background system should by definition reflect the wider technological and economic developments. This requires narratives and models that depict possible futures of entire sectors, such as energy generation, raw materials supply, manufacturing, or waste treatment, and ultimately, the creation of an LCI database that reflects these scenarios. Future background scenarios have been developed previously, notably the NEEDS database (New Energy Externalities Development for Sustainability) (NEEDS 2009) and the THEMIS model (Technology Hybridized Environmental-Economic Model With Integrated Scenarios) (Gibon et al. 2015; Hertwich et al. 2015). More recently, Mendoza Beltran et al. (2018) have presented future background scenarios that result from a combination of the ecoinvent database (Wernet et al. 2016) and the integrated assessment model IMAGE (Stehfest et al. 2014). Integrated assessment models (IAMs) are broad models that represent global economic, technological, and social processes and their interactions with the environment, e.g. the climate system (Moss et al. 2010). IAMs typically implement globally consistent future scenarios that cover plausible evolutions of society and ecosystems over a century timescale (O'Neill et al. 2014), such as the Shared Socio-Economic Pathways (SSPs) (O'Neill et al. 2014; van Vuuren et al. 2014).

The combination of IAMs with LCI databases and specifically the work of Mendoza Beltran et al. (2018) is interesting for two reasons: first, it brings already well-established future scenarios to the LCA community and thus avoids the need to re-invent these for LCA. Second, the large scale systematic modifications of an LCI database in Mendoza Beltran et al. (2018) have been based on a dedicated python package for systematic database modifications called wurst (Mutel 2020). A code-based rather than a manual

way to generate *scenario LCI databases*[1] is an important step towards a stronger and more permanent integration of information from LCI databases and IAMs. It also improves reproducibility and transparency and reduces the effort to generate updates when new versions of the underlying models become available. This work is currently being continued in the context of the PREMISE,[2] a python package that aims at streamlining the approach to produce scenario databases for prospective LCA (Sacchi et al. submitted), and which has already been applied in a number of studies (Pizzol et al. 2021; Sacchi et al. 2021). Despite this progress, there are still important challenges to be met for enabling a more widespread use of future background scenarios in LCA (see also our discussion).

Here, we address one of these challenges, which is a key technical problem that currently limits the practical usability of background scenarios. The problem is that such scenarios have until now been generated as individual LCI databases. For example, each scenario and reference year in Mendoza Beltran et al. (2018) is generated as a separate scenario LCI database (e.g. 6 scenarios × 4 reference years = 24 databases, each of which contains roughly 15′000 processes). These scenario databases can be imported and used in LCA software like any other LCI database. However, there are two issues with this approach. The first one concerns the linking of foreground systems (i.e. processes modelled by the practitioner) with future background systems. For example, a practitioner may model the production of a chemical, which requires electricity and for which a process from a scenario database is used. If he then wishes to assess his foreground system against another scenario or reference year, all inputs from the background system to the foreground system need to be replaced by inputs from another scenario database (as illustrated in Fig. 1, left side). This is very impractical and essentially a game-stopper for analyzing, e.g. novel technologies against different future scenarios and reference years. The second issue relates to the quantity of data stored: when only selected parts of the future background databases differ across scenarios (e.g. electricity generation technology and market shares, as described by Mendoza Beltran et al. (2018)), parts of the future background databases are identical in all scenarios. This means that potentially large amounts of duplicate data are stored across the individual LCI databases, which negatively affects the required hard disk space and the speed of LCA calculations as the same data is loaded several times.

---

[1] We define the term "scenario LCI database" here as an LCI database that represents a specific scenario and reference year (e.g. a "business as usual" scenario for 2030 or a "sustainable development" scenario for 2040). Throughout this paper, we often abbreviate the term to "scenario database" to make it shorter.

[2] PRospective EnvironMental Impact AsSEssment (PREMISE). https://github.com/romainsacchi/premise
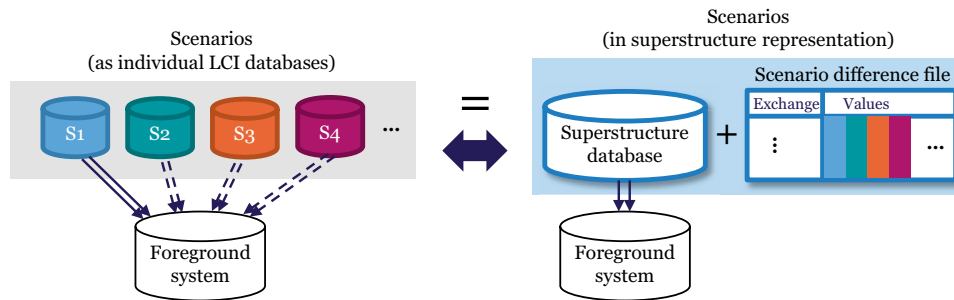
Fig. 1 Concept of the superstructure approach: any number of scenario LCI databases can be represented through a superstructure database and a scenario difference file that stores the data that differs between the scenarios. Each representation can be translated into the other; i.e. the superstructure and scenario difference file can be created from a number of scenario LCI databases and, vice-versa, scenario LCI databases can be created from the superstructure with the help of the scenario difference file. When using individual scenario LCI data- bases to represent different scenarios (left side), the inputs from one scenario database need to be replaced with inputs from another scenario database, if the practitioner wishes to assess his foreground system against a different scenario. A key advantage of the superstructure approach (right side) is that the same inputs from the superstructure database to the foreground system can be maintained, while the superstructure database can be modified based on the scenario difference file to represent different background scenarios

In this paper, we propose a solution that we call the superstructure approach, which can help to solve the linking problem and to minimize the data quantity problem. In the following section, we introduce the approach and illustrate it at the example of a small case study implemented in the Activity Browser open source LCA software (Steubing et al. 2020). Finally, we discuss its limitations, requirements for LCA software, and further challenges for making the use of future background scenarios in LCA more widespread.

2 Superstructure approach

2.1 Concept and definitions

A life cycle inventory (LCI) database describes the flows of products between processes of an economic system (intermediate flows) and the interaction of the economic system with the environment (elementary flows). The concept behind the superstructure approach presented here is that any number of individual LCI databases (scenarios) can be represented through a single "superstructure" LCI database and additional data that specifies the differences between the scenarios (thus called scenario difference file; see Fig. 1). This is possible if the superstructure database contains all unique processes and flows that occur across all scenario databases and if the scenario difference file contains all flow values that change between scenarios. A superstructure database shall thus be defined as an LCI database that contains all possible economic structures and their interactions with the environment of a set of individual LCI databases (i.e. the scenario databases). The scenario difference file shall be defined as a file that stores the flow values for all flows that differ between the individual LCI databases (scenarios). The superstructure database and the scenario difference file can be generated from the individual

scenario databases. Vice versa, any scenario database can be reconstructed by applying the data for a specific scenario from the scenario difference file to the superstructure. The individual scenario databases and the superstructure plus scenario difference file are thus merely different representations of the same data.

However, there are two *advantages* of representing future background scenarios using the superstructure database approach: first, the *LCA practitioner can work with a single background database* (the superstructure) instead of having to work with a number of scenario databases (see Fig. 1). This greatly simplifies the modelling process as foreground processes do not have to be re-linked to different background databases when performing LCA calculations for different scenarios. Second, the superstructure representation is likely to be much *more compact* since only one full LCI database is generated (the superstructure), while for all scenarios, only the differences to the superstructure database are stored in the scenario difference file. Hence, data that is not changing across scenarios is only stored in the superstructure (no duplication).

2.2 Generation of superstructure database and scenario difference file

2.2.1 Illustrative example

To explain how the superstructure approach works, let us introduce two slightly differing scenario databases as shown in Fig. 2. In the first scenario, there are two processes: (1) a natural gas power plant that supplies and (2) the market for electricity. The second scenario is similar to the first scenario, with the following differences: a third process supplies electricity from a wind turbine to the market for electricity and replaces half of the electricity that is supplied by natural gas in scenario 1. Additionally, scenario 2 reflects technological improvements
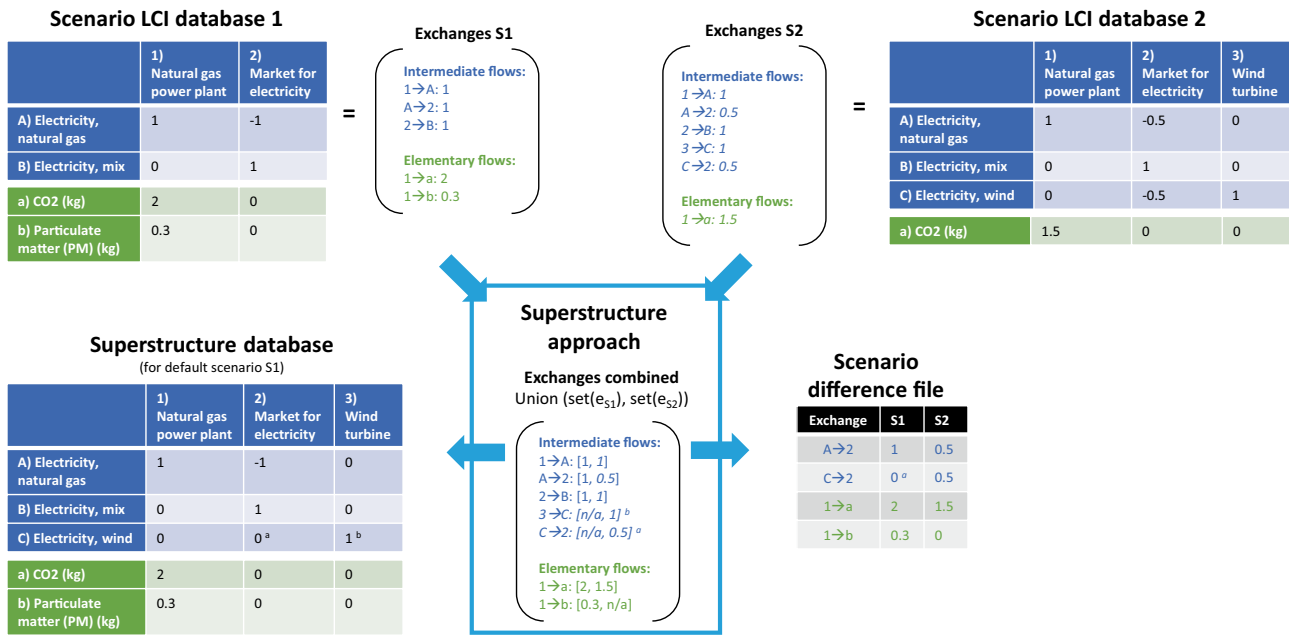
Fig. 2 Example for the generation of a superstructure database and the corresponding scenario difference file from two scenario LCI databases that differ both in economic structure and flow values. The superstructure database is shown for the default scenario 1. When the values from scenario 2 in the scenario difference file are applied to the superstructure, the matrix becomes identical to the matrix of scenario 2. Blue tables represent the technology matrices with inter-mediate flows; green tables the intervention matrices with elementary flows; [a]an exchange that is not part of all scenarios describing an input to a process (it is set to 0 in all scenarios that it is not part of); [b]an exchange that is not part of all scenarios that describes the function of a process (in this case its output; its value is set to 1 in scenarios that it is not part of in order to avoid empty columns in the technology matrix, which would break LCA calculations)

for the natural gas power plant, which now emits less CO_2 per unit of electricity due to efficiency increases and particulate matter emissions are reduced to zero (all numbers are invented and serve only for illustration purposes).

By applying the superstructure approach, a superstructure database can be generated that includes all three processes and all intermediate and environmental flows. The superstructure database is shown with values for the arbitrarily chosen default scenario 1. Note that the wind turbine is included here, but contains no inputs and is not used by any other process and thus does not interfere with other products (reasons for this are discussed in Sects. 2.2.5 and 4.3). The scenario difference file contains the exchange values that differ between scenarios. In this example, only 4 exchange values need to be changed to turn the superstructure database into a database representing scenario 2.

2.2.2 Representing LCI databases as lists of exchanges

Note that in Fig. 2 we use both a matrix and an exchange representation of the scenario databases. Traditionally, intermediate flows between economic processes are described in the technology matrix **A** and elementary flows between the economic system and the environment are described in the interventions matrix **B** (Heijungs and Suh 2002). The same

information can also be described by a list of intermediate and elementary flows. As an umbrella term for both flow types, we use "exchange" and "exchange value" for the flow value. An example for an intermediate flow could be "1 kWh of [A] electricity, (Arvidsson et al. 2018) natural gas from a natural gas power plant to Cox et al. (2020) the market for electricity" or, in an abbreviated notation, "A ➔2: 1.0" (see Fig. 2). An example for an elementary flow could be "2 kg [a] CO_2 from a (Arvidsson et al. 2018) natural gas power plant to the environment", or in an abbreviated notation, "1 ➔a: 2.0." Note that in our notation, processes are abbreviated by numbers; intermediate flows are abbreviated by capital letters and refer to products from a specific supplier (e.g. we know that product A is produced by process 1), elementary flows are represented by lowercase letters, and exchange values are given after the colon. Note also that in matrix notation, inputs are negative and outputs are positive numbers, while in the exchange notation, the directionality is given by the arrow.

2.2.3 Identifying differences between scenarios

Describing scenario LCI databases as lists of exchanges helps to understand the differences between scenarios. There are two ways in which exchanges of different scenario databases can differ (see example in Fig. 2):

Table 1 Possible cases for exchange values and respective solutions to store this data in the superstructure approach (SDF: scenario difference file)

Case	Exchange present in all scenarios?	Exchange characteristic	Data stored in the superstructure	Data stored in the scenario difference file
1	Yes	Same value in all scenarios	Value of default scenario	-
2		different values across scenarios	Value of default scenario	Value for each scenario
3a	No	Exchange describes a flow between two activities (e.g. "C➔2" in Fig. 2)	Value of the default scenario, if this exchange is present in the default scenario, otherwise "0"	Value for each scenario, or "0" if not present in a specific scenario
3b		Exchange describes the function of a process (e.g. its output; e.g. "3➔C" in Fig. 2)	Value of the default scenario, if this exchange is present in the default scenario, otherwise "1"[1]	Value for each scenario, or "1"[1] if not present in a specific scenario

[1]In principle, also, this exchange value should be set to "0" in scenarios where this exchange is not present (meaning that these scenarios do not contain the associated process). Please see our discussion on this choice in 2.2.5 and 4.3

1. The *exchanges can be different*, meaning that the structure of the economic system or the interactions with the environment differ across scenarios (e.g. the input of wind power to the market for electricity "C ➔ 2" exists only in scenario 2), and
2. The *exchange values can be different*, meaning that the magnitude of intermediate or elementary flows differ across scenarios (e.g. the amount of natural gas electricity input to the market for electricity is "A ➔ 2: 1.0" in scenario 1 and "A ➔ 2: 0.5" in scenario 2).

2.2.4 Obtaining the superstructure

By definition, the superstructure database shall contain all unique elementary and intermediate flows (exchanges) from all scenario databases (or, when viewed as matrices, all unique processes (columns) and flows (rows) of all scenario A and B matrices). In mathematical terms, all unique exchanges can be obtained by a union of the sets of exchanges *e* of all scenario LCI databases, as in Eq. (1),

$$e_{superstructure} = \bigcup_{i=1}^{n} e_i \tag{1}$$

where *i* represents the *i*th scenario database and *n* is the total number of scenario databases. Each set of exchanges consists of all intermediate and elementary flows of a given scenario database (as illustrated in Fig. 2). The economic structure and the interactions with the environment of each scenario database are thus fully represented by a subset of $e_{superstructure}$. The exchange (or matrix) values need to be discussed next.

2.2.5 How to store exchange values

Scenario data can be stored both in the superstructure database and in the scenario difference file (SDF). For the scenario difference file, a spreadsheet is well suited. Each row

in this spreadsheet consists of two parts: (a) information necessary to identify an exchange in the superstructure and (b) the values for all scenarios (see also Fig. 2 for the concept and Table 2 for an implementation of a SDF). In order to determine where to store exchange values, several cases need to be distinguished (Table 1):

Case 1: if an exchange is present in all scenarios and its value does not differ across scenarios, the exchange value can be directly stored in the superstructure database and there is no need to store data related to this exchange in the SDF.

Case 2: if an exchange is present in all scenarios, but its value differs, the exchange values of all scenarios are recorded in the SDF. Additionally, the value from a chosen default scenario can be stored in the superstructure database. Since case 2 values in the superstructure are intended to be overwritten by values in the SDF, this is not strictly necessary. Yet, doing so provides the possibility of using the superstructure database without the SDF to represent a default scenario (thus, if the LCA practitioner is interested in LCA for the default scenario, there is no need to apply any data from the SDF to the superstructure database, while all other scenarios can be obtained by applying data from the SDF to the superstructure).

Case 3a: if an exchange is not present in all scenarios, exchange values for all scenarios are recorded in the SDF. For scenarios where the specific exchange is not present the value is set to "zero." Additionally, as in case 2, the exchange value of a default scenario can be stored in the superstructure, or zero, if the exchange is not present in the default scenario.

Case 3b: A special case is exchanges that describe the function of a process (typically its output; in Fig. 2 the diagonal values in the A matrices). Technically, these should be treated like in case 3a, i.e. values set to zero in

Table 2 Structure of the scenario difference file that is created from our superstructure python library (https://github.com/LCA-ActivityBrowser/brightway-superstructure), which was used for our case study. Note that the table contains two "key" columns, which can be used alternatively to the other columns to reference the to/from parts of an exchange by an identifier

| | Part a) Exchange identification | | | | | | | | | | | | Part b) Exchange values | | | | |
| | Flow from … | | | | | | … to | | | | | | Scenarios | | | | |
	from activity name	from reference product	from location	from categories	from database	from key	to activity name	to reference product	to location	to categories	to database	to key	SSP2-base, 2020	SSP2-2.6, 2020	…	SSP2-base, 2050	SSP2-2.6, 2050
example intermediate flow	market for natural gas, high pressure	natural gas, high pressure	US		ei37_cutoff_IMAGE_SSP2_SS		electricity production, natural gas, combined cycle power plant	electricity, high voltage	US-NPCC		ei37_cutoff_IMAGE_SSP2_SS		0.183	0.182	…	0.150	0.147
example elementary flow	Carbon dioxide, fossil			('air', 'non-urban air or from high stacks')	biosphere3		electricity production, natural gas, combined cycle power plant	electricity, high voltage	US-NPCC		ei37_cutoff_IMAGE_SSP2_SS		0.381	0.380	…	0.312	0.306

the SDF and superstructure database for scenarios where these exchanges are not present. This would effectively lead to an exclusion of processes that are not present in a certain scenario. Although this would conceptually and technically be the right choice, we decided to set these values to "one" in the SDF and superstructure database for scenarios where these exchanges are not present. As a result of this convention, all processes are included in the superstructure and scenario databases derived from the superstructure and SDF (e.g. the process "3) (Wind turbine" in Fig. 2). However, in the scenarios where they technically should not exist, they are not connected to any other process and thus do not contribute to LCA results (i.e. they will have no environmental impacts associated with them). We still opt for this choice as it has certain advantages: all processes and flows across all scenario databases are already included in the superstructure database. Therefore, when doing LCA calculations, the A and B matrices can be constructed from the superstructure and turning the superstructure into a scenario database via the SDF only requires changing values in the matrix, but not its structure. Further, setting case 3b values to "zero" translates columns of all zeros in the A and B matrices for processes that not present in a given scenario. This could pose a problem and break LCA calculations in LCA software that does not automatically detect and remove such columns (see also our discussion on this choice in Sect. 4.3).

2.3 Workflow for using the superstructure approach

Figure 3 proposes a generic workflow for using the superstructure approach within the context of prospective LCA studies. It distinguishes 4 phases. In the first phase, the scenarios are generated and, ultimately translated to scenario LCI databases (step 1). This may in reality be a large and complex process that involves, for example, the generation of narratives (such as the SSP scenarios), the representation of such narratives in quantitative models (such as IAMs), and finally the mapping of different data sources to existing LCI databases, as described for example by Mendoza Beltran et al. (2018). Note that in contrast to Mendoza Beltran et al. (2018), we include the translation of data from different data sources into LCI data as part of scenario generation as we believe that additional assumptions still need to be made at this level, although one could argue that this is not scenario generation anymore, but merely a "translation" of data from one model to another. In the second phase (step 2), the individual LCI databases are converted to a superstructure database and scenario difference file using the superstructure database approach.

In the third phase, this data is shared with LCA practitioners. There are various conceivable ways of sharing this data, such as (a) sharing the actual superstructure database and scenario difference file (e.g. through an online platform where it can be downloaded) or (b) providing a software tool that permits the LCA practitioner to generate the superstructure database and scenario difference file locally, such as PREMISE (Sacchi et al. submitted). Both solutions will have to consider potentially licensed data (e.g. the ecoinvent database). For (a) this could be solved by only allowing users with a license to download the superstructure data, while for (b) this could be solved by requiring the user to have the licensed data on his own computer.

In the fourth phase, the LCA practitioner is in possession of a superstructure database and scenario difference file and wants to use future backgrounds for prospective LCA. We include more steps here as this phase describes the workflow that LCA practitioners will be most concerned with. The steps that LCA practitioners need to follow are as follows: step 3: to import the superstructure database; optional step 4: to perform any additional modelling that is required for a specific prospective LCA study and link such a foreground system to the background system represented by the superstructure database; step 5: to setup the scenario LCA calculations, i.e. to define functional units, impact categories and
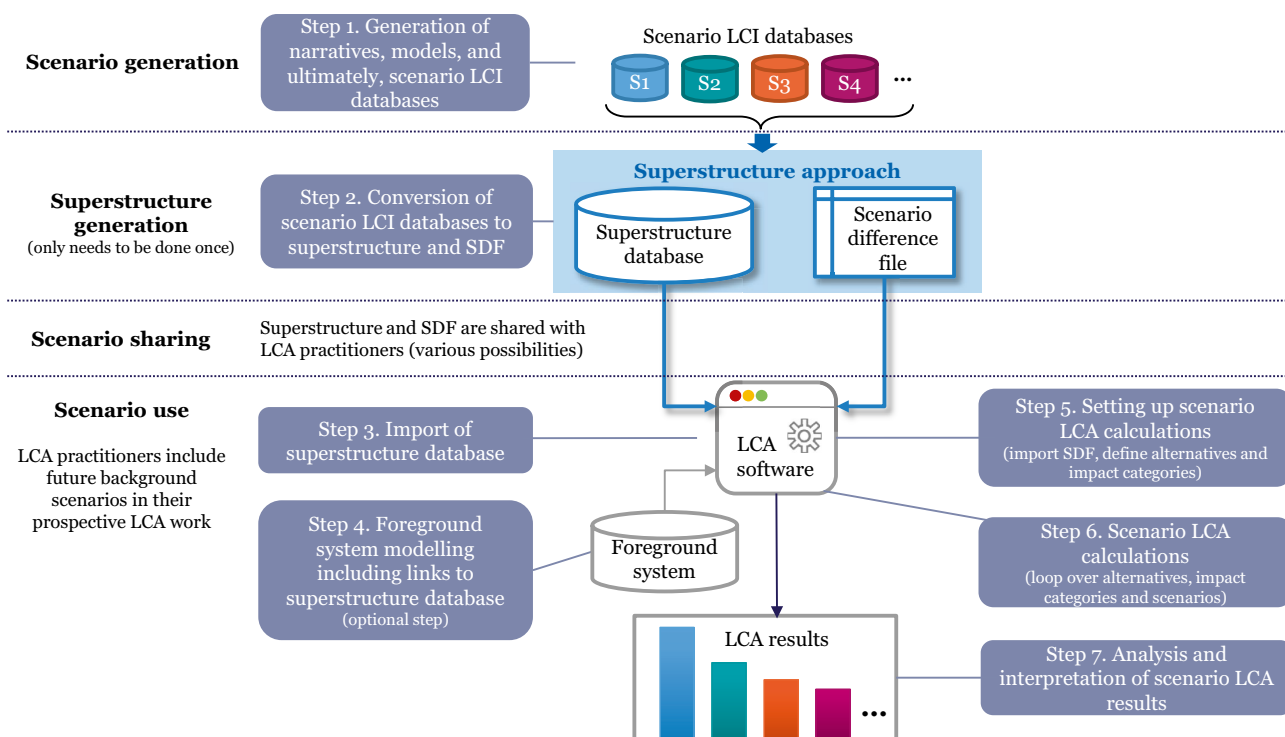
Fig. 3 Generic workflow for using the superstructure approach (SDF = scenario difference file)

scenarios to be analyzed; step 6: to perform scenario LCA calculations; and step 7: to analyze the scenario LCA results.

3 Case study and software implementation

To illustrate the practical application of the superstructure approach and the importance of future background data, we present a simple case study for electric vehicles using future background scenarios for from 2020 to 2050. In the following, we describe the modelling of this case study according to the steps of the generic workflow outlined in Fig. 3.

Step 1: Generation of narratives, models, and ultimately, scenario LCI databases. The scenario databases are derived from a combination of the ecoinvent 3.7 (cut-off system model) and the IMAGE 3.0 database using the python notebooks provided by Mendoza Beltran et al. (2018). Small adaptations were made to make the data and code compatible with the ecoinvent database version 3.7. We include two main scenarios, the Middle of the Road base-scenario (SSP2-base) that follows a representative concentration pathway (RCP) of 6 W/m^2, and a more ambitious Middle of the Road scenario that follows RCP 2.6 (SSP2-2.6). For both scenarios, we generate four scenario databases representing the years 2020, 2030, 2040, and 2050, which leads in total to 8 scenario databases.

We do not describe here the generation of the SSP scenarios and the IMAGE model as these steps are already well documented in the literature (e.g. O'Neill et al. 2014; Stehfest et al. 2014).

Step 2: Conversion of scenario LCI databases to superstructure and SDF. The superstructure approach as described in the method section is used to convert the 8 scenario databases into a single superstructure database and a scenario difference file. The code to generate superstructure databases and corresponding scenario difference files from scenario databases is provided on GitHub[3] and builds upon the brightway LCA framework (Mutel 2017). Table 2 shows the structure of the SDF, which consists of information to identify exchanges and the exchange values in different scenarios. Intermediate flows are uniquely identified by the activity name, reference product name, location, unit, and database that is referred to for both the supplying and the receiving process. Elementary flows are uniquely identified by a "from part" that consists of the name of the elementary flow, category information, and the database it is stored in, while the "to part" identifies the process that is responsible for the elementary flow. Note that the convention used here for elementary flows uses the "to" and "from" in opposite direction as defined

---

[3] https://github.com/LCA-ActivityBrowser/brightway-superstructure

in the "Method" section. This is due to the implementation in brightway, where the from part of an exchange describes the elementary flow and the to part the process that is responsible for this flow. The directionality is then based on category information (e.g. "air" indicates that this is an output of a process).

Step 3: Import of superstructure database. The superstructure database was imported into the Activity Browser open source LCA software (Steubing et al. 2020), which builds upon the brightway LCA framework (Mutel 2017).

Step 4: Foreground system modelling including links to superstructure database (optional step). A foreground system was modelled consisting of two processes: first, a copy of the ecoinvent unit process "transport, passenger car, electric (GLO)" and second, an improved electric vehicle (EV). The improvements are supposed to reflect the use of lighter materials, higher drive train efficiency, and improved batteries with a higher energy density. For simplicity, we assumed that such improvements result in 40% lower electricity consumption and a 40% smaller battery (numbers are invented and for illustrative purposes only). A second copy of the EV process in ecoinvent was thus made, and the electricity and battery inputs were reduced each by 40%. We then replaced all inputs to our two EV foreground processes from the ecoinvent database with equivalent inputs to the superstructure database using an automatic re-linking function available in the Activity Browser.[4] This function first identifies all exchanges between the foreground system with a given LCI database (e.g. ecoinvent) and then tries to replace these exchanges with equivalent exchanges (i.e. with the same product, activity, location, and unit name) in another LCI database (here the superstructure).

Step 5: Setting up scenario LCA calculations. A calculation setup was created with the reference flows of "driving 1 km" with each electric vehicle. Life cycle impact assessment (LCIA) is performed using the climate change (100-year time horizon) indicator (IPCC 2013). The SSP2-base and the SSP2-2.6 scenarios were assessed for all reference years, as shown in Fig. 4.

Steps 6: Scenario LCA calculations. We used the Activity Browser to automatically calculate the LCA results for all alternatives, impact categories, and scenarios. LCA calculations are performed by three nested *for* loops. The innermost loop iterates over the impact categories (here only one). The middle loop iterates over the reference flows. The outer loop iterates over the scenarios and changes the data of the A and B matrices in memory for each scenario and reference year based on the data specified in the scenario difference file. Approximately 135,000 exchange values are overwritten in the A and B matrices for each scenario and reference year. The calculation time on our laptop was 5 s for the 16 LCA calculations (2 scenarios, 4 reference years, 2 reference flows, and one impact category).

Step 7: Analysis and interpretation of scenario LCA results. We also used the Activity Browser for the analysis of the LCA results. The consideration of scenarios adds an additional dimension to the analysis of LCA results (on top of reference flows and impact categories), and LCA software thus may need to implement additional functionality for analyzing and comparing scenario LCA results, e.g. as shown in Fig. 5.

Figure 6 shows a contribution analysis for the LCA results of both EVs over time in the two scenarios. It can be observed that electricity generation is the single biggest contributor to climate impacts. The improved EV always performs better than the regular EV in our comparison, as it is virtually identical, except that it consumes 40% less electricity. Although this leads to substantial GHG reductions, the energy transition has a bigger leverage to reduce climate impacts than the efficiency improvement alone, as shown in the SSP2-2.6 scenario. Also, the advantage of the improved EV over the regular EV shrinks with the progression of the energy transition due to the decreasing relative importance of the power production sector for climate impacts. Obviously, the energy transition and energy efficiency work towards the same goal and should both be pursued. These findings are in line with observations by Mendoza Beltran et al. (2018) and confirm the importance of using future background data as the influence of the background system on LCA results can be very significant. Thus, not considering future developments of the background system may lead to the drawing of sub-optimal conclusions (obviously, it depends on the technology that is being assessed how large the influence of the background is).

4 Discussion

4.1 Contribution of this work

Recently, great progress has been made in generating future background scenarios for LCA based on a combination of existing LCI databases and data from other models that represent future societal and technological developments, such

---

[4] "Re-link database" is a simple function in Activity Browser that tries to replace all inputs in a given database from a specific database with inputs from another database. For example, the database "foreground system" may contain inputs from the database "ecoinvent." The re-linking function could then be used to iterate over all processes in the "foreground system" database and replace inputs from the "ecoinvent" database with equivalent inputs from a "superstructure" database. Inputs are replaced, whenever product, process, unit, and geography names match, for example, "electricity, high voltage | market for electricity | kWh | NL | *ecoinvent*" would be replaced with "electricity, high voltage | market for electricity | kWh | NL | *superstructure*.".

Fig. 4 Screenshot of the Activity Browser software showing the three parts of a "Scenario LCA" calculation setup: definition of reference flows, impact categories, and scenarios. The scenario difference file is loaded and provides the data that differs between the background scenarios and reference years

as IAMs. While these databases can be directly used in LCA software, this is not ideal from a modelling perspective. The superstructure approach presented in this paper proposes an intermediate step, as shown in Fig. 7, where the individual scenario databases are converted to a superstructure database and a scenario difference file, before the future backgrounds are used in LCA software. This *solves two problems* that are associated with the practical use of future scenarios in LCA:

- First, it provides a solution for the *linking problem* that arises when alternative background databases are introduced. The solution is the superstructure database itself to which a foreground system can be linked and since the superstructure can be modified based on the scenario difference file to represent different scenarios, the links to

the background database can be permanent. This makes it much easier to use future background scenarios in practice. It further opens up possibilities for *fast and automated LCA calculations for all scenarios*. A key element here is that the A and B matrices need to be constructed only once from the superstructure, which then represents all possible economic structures and their interactions with the environment. All LCA software can already do this; the only additional step is then a for loop in the LCA calculation during which only those matrix values that change across the scenarios in the SDF are overwritten. This can be done in memory and is, therefore, very fast.

- Second, the superstructure approach stores, in principle, *no duplicate data* and, therefore, requires significantly less hard disk space, especially in situations where sce-
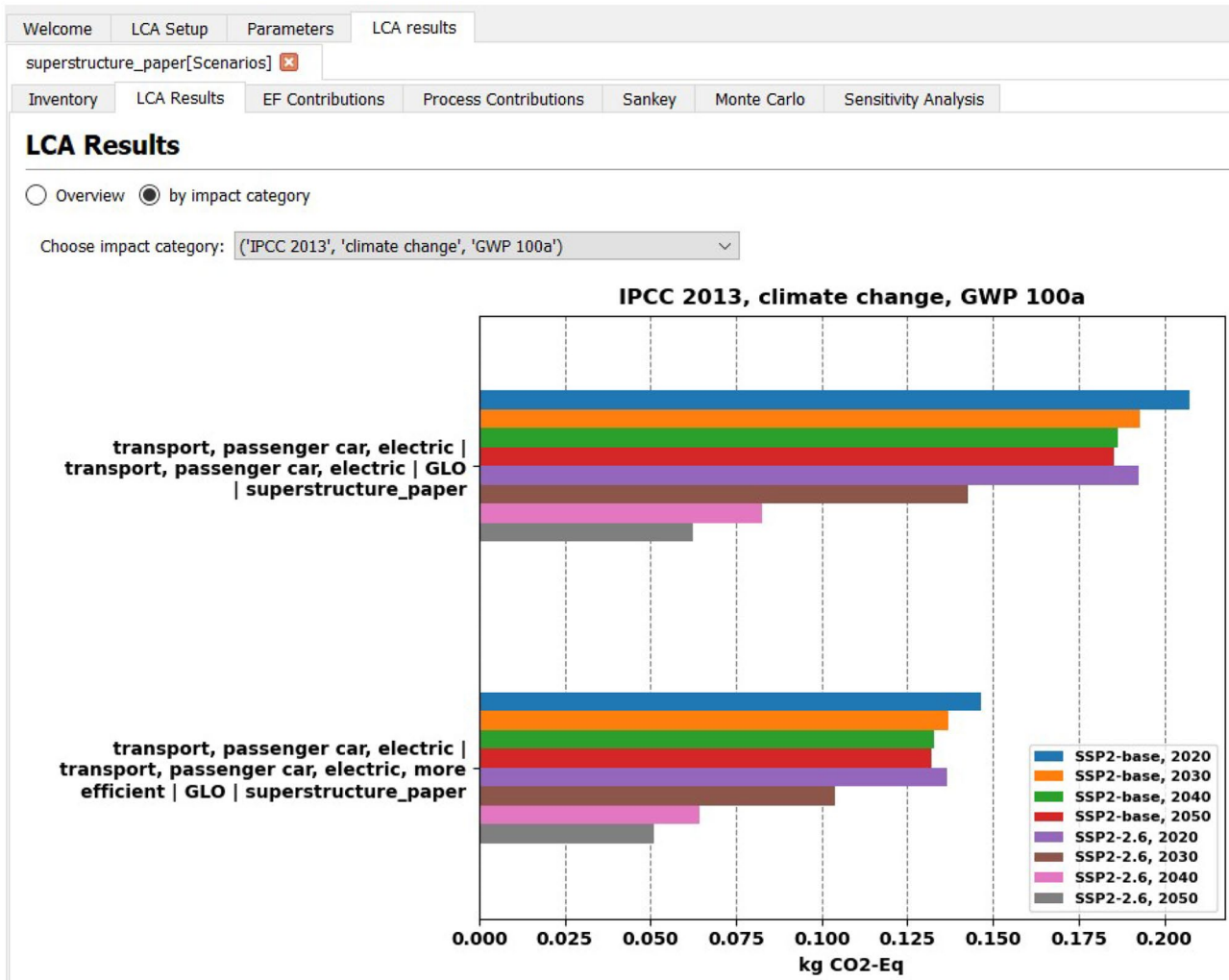
Fig. 5 Scenario LCA results in the Activity Browser. Note that the inclusion of scenarios adds an additional dimension to LCA results (on top of reference flows and impact categories), for which functionality is available in the Activity Browser, e.g. to compare process contributions by scenario for a specific reference flow and impact category

nario data is only translated to certain parts of an LCI database (e.g. the energy sector). However, even if every exchange value was to change across scenarios, the superstructure approach would still require considerably less hard disk space, since the SDF, at least in our implementation, does not store any metadata, which make up the bulk of the data in an LCI database (general metadata can still be stored in the superstructure).

The modification of exchange values during LCA calculations based on values specified in the scenario difference file also opens up further opportunities, such as the following:

- *Practitioners can easily modify or extend existing scenarios* by changing the values in the SDF or by adding additional scenario data for specific sectors. Examples for such scenarios are the recent work on future

metal supply scenarios by Harpprecht et al. (2021) and Meide et al. (submitted), which could be added to the scenarios of Mendoza Beltran et al. (2018) to combine energy and metal scenarios.

- *The SDF provides a generic and powerful tool to for scenario modelling in both the foreground and background systems.* Although the focus of this paper has been on future background systems, the SDF can be used to specify alternative values for any exchange and is, therefore, not limited to background LCI databases. The Activity Browser introduces the possibility of using several SDFs simultaneously, e.g. one for the background system and one for the foreground system, as shown in Fig. 7. The software also allows for a combinatorial use of SDFs, which enables practitioners to analyze combinations of scenarios (e.g. each foreground scenario against each background scenario).
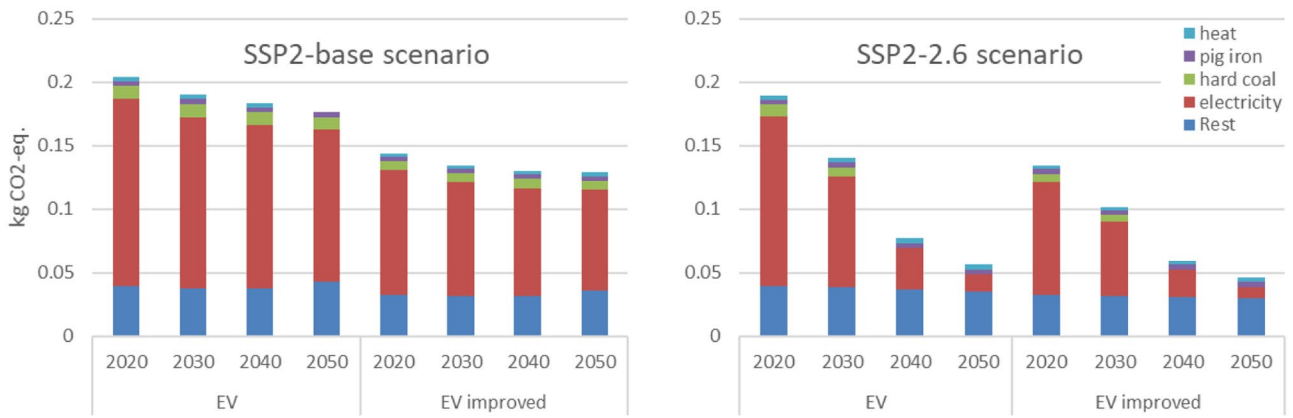
Fig. 6 Climate change impacts for driving 1 km with an electric vehicle and an improved electric vehicle over time within the SSP2-base and the SSP2-2.6 scenarios. Numbers for the improved electric vehicle are based on assumptions that were made purely for illustrative purposes, assuming a 40% smaller battery and a 40% reduced electricity consumption

- *The SDF facilitates sharing and peer-reviewing.* The SDF as a simple spreadsheet is human readable and editable with standard office software. This, and the likelihood that the data contained in the SDF is not under license (e.g. it is not from the ecoinvent database, but from an IAM or from an LCA practitioner), facilitates sharing and peer reviewing of scenario data.
- *The SDF could also be used to represent stochastic data* in the sense that each column could represent one set of values from a stochastic model. This is similar to the idea developed in the presample framework (Lesage et al.

2018), which has been an important source of inspiration for the superstructure approach. However, when large numbers of scenarios (or samples) need to be evaluated, other data formats may be computationally more efficient.

4.2 Requirements for LCA software

The implementation in the Activity Browser has yielded the proof of concept that the superstructure approach can be practically implemented in LCA software in a meaningful
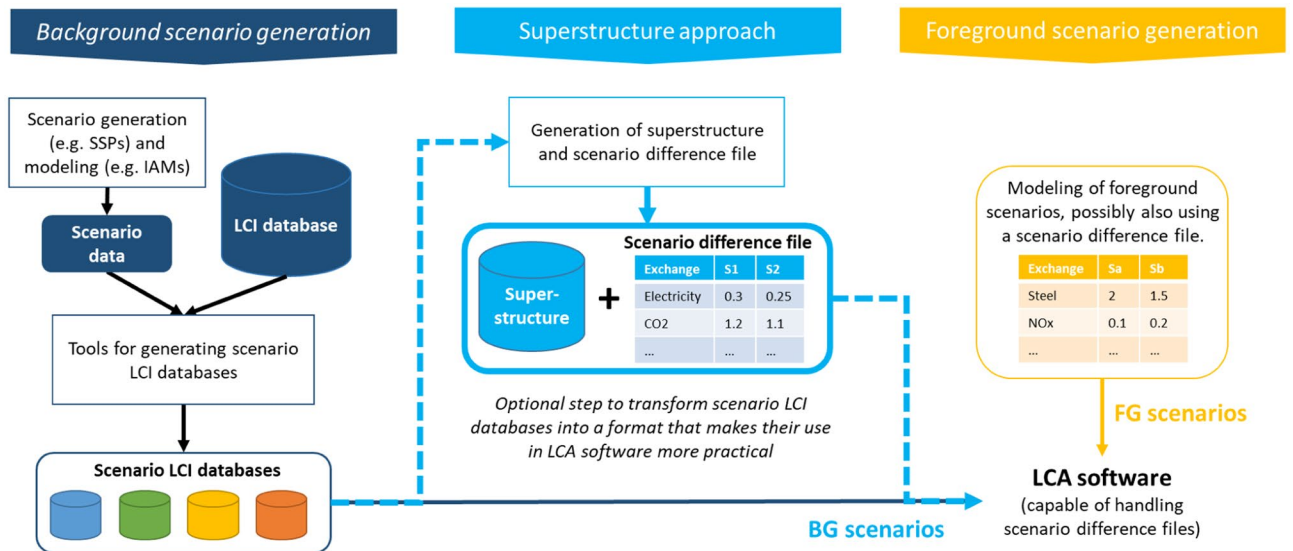

Fig. 7 The superstructure approach is an optional, but useful step to transform individual scenario LCI databases to a single LCI database (the superstructure) and an associated scenario difference file. This supports the use of scenario LCI databases in LCA practice as it overcomes the need to re-link foreground systems to multiple background databases and thus facilitates the evaluation of products and services against different background scenarios. The scenario difference file is a general purpose file that can also be used for modelling foreground scenarios

way to improve the way we can use future background scenarios. The authors hope that the superstructure approach, or a variant of it, can also find its way into other LCA software. Some of the functionality that would have to be added concerns the following:

1. Import of scenario difference files: Functionality will have to be added to LCA software to import a scenario difference file. We suggest a spreadsheet format as provided in Table 2 to have a file that can be easily read and modified, if desired, by LCA practitioners.
2. *Scenario LCA calculation:* We believe that it is most useful for LCA practitioners, if the software is able to calculate LCA results for several scenarios at once. This leads to several requirements: the user needs to be able to specify which scenarios shall be included; the software should then iterate over all included scenarios and do LCA calculations for each scenario; for each calculation, the superstructure will have to be modified using the data for a specific scenario from the scenario difference file, which is, in our opinion, best handled in memory and not written to disk to improve speed; calculating LCA results for a number of scenarios adds an additional dimension of data (in addition to LCA results data for, e.g. different functional units and impact categories), and thus, the data format for storing LCA results may have to be extended accordingly.
3. *Scenario LCA results analysis:* Functionality will have to be added to choose between or compare scenarios when analyzing LCA results at different levels (inventory, characterization, contributions, etc.).

4.3 Limitations

Although the superstructure approach represents in our opinion a milestone towards a more widespread use of future background scenarios in LCA, certain aspects may have to be improved or revised in future implementations.

We initially describe a set of individual databases and the superstructure plus scenario difference file as mathematically equivalent representations of the same data (Fig. 1). However, we then slightly deviate from this concept by including all processes, i.e. also processes that are only present in certain scenarios (case 3b in Table 1), in the superstructure database. The disadvantages of this choice are that the superstructure may contain ghost-like processes that deliver a product or service without having any process inputs or environmental flows (although data for such processes is added via the SDF in scenarios where these processes are meant to exist) and that a reconstruction of individual scenario LCI databases from the superstructure database and SDF would contain such ghost processes. As shown in our example in Fig. 2, the

wind turbine is such a process. It is not part of scenario 1, but still it is included in the superstructure, albeit with no process inputs and no other process using it when applying SDF values for scenario 1. Scenario generators should document such processes when providing the scenario databases in the superstructure format to make LCA practitioners aware that certain processes should only be used in the context of selected scenarios. While the superstructure approach could easily be adapted to include processes only in scenarios where they are meant to be present (by treating case 3b like case 3a in Table 1), we have opted for this choice due to the simple logic it follows: keeping all processes in the superstructure database avoids the need to include additional processes (that occur only in selected scenarios) for individual scenarios during LCA calculations. Thus, the A and B matrices only need to be constructed once from the superstructure and then the iteration over all scenarios involves nothing but the change of selected values in the matrices based on the SDF (instead of the construction of new matrices). This approach may also be the easiest to adopt by other LCA software.

In the method section, we mention both matrix and exchange notation of elementary and intermediate flows. We used the exchange notation as a more intuitive way to explain how the set union of exchanges from all scenario databases can yield all unique exchanges and thus the superstructure. Mathematically, an LCI database is a graph and we can learn from graph theory that graphs can be represented both as matrices or as edge lists (in LCA terminology the exchanges). However, there is currently, despite first attempts (Heijungs 2015), no rigorous mathematical treatment of LCA using graph theory. Such treatment would be helpful to properly describe work like ours using a graph theoretical notation.

We have not discussed uncertainties in this paper. While it could be disputed whether it is meaningful at all to use uncertainty data for future scenarios that look decades ahead, we have also not considered the possibility for including uncertainties for scenario data in the SDF and leave this for future research.

LCI databases are typically rich in metadata that describe specific modelling choices and data sources at the process or flow level. One of the potential drawbacks of the superstructure approach over a representation of future scenarios in separate LCI databases is that it is unclear where the documentation of modelling choices at the level of the individual scenarios should be stored. Technically, additional metadata could be stored in the superstructure database, the SDF, or in another place. Yet, this relates to the perhaps bigger question of how future background scenarios for LCA should be documented in general.

Finally, while the superstructure approach makes it easier to use scenarios in prospective LCA, it is a technical solution

only and no guarantee for the meaningfulness and quality of the scenarios that it can represent. LCA practitioners should thus not blindly use future scenarios provided in this format, but instead attempt to understand in sufficient depth what the scenarios represent and how they are modelled in order to avoid wrong conclusions.

4.4 Challenges for a wider use of background scenarios in LCA

Despite the recent progress in generating future scenarios for LCA and the technical solutions presented here, there are still a number of important questions and challenges to be addressed for enabling a more widespread use of future scenario databases (see also (Vandepaer and Gibon 2018)). Some of these relate to the following:

– *Scenario generation:* The generation of scenario LCI databases is typically a further step in a chain of models, including, e.g. IAMs and LCI databases, and each model comes with specific paradigms, assumptions, and limitations. Scenario LCI database generators need to carefully consider which data sources can be combined to generate consistent and state-of-the-art future scenarios (promising work is currently ongoing in the context of PREMISE (Sacchi et al. submitted)). When generating scenario LCI databases for different underlying LCI databases (or different system models of ecoinvent), a key challenge are differences in how economic sectors are represented in these databases, as well as differences in the naming of elementary and intermediate flows (lack of harmonization). Ultimately, we believe that it would be most useful for the LCA community to use a small but diverse set of well-accepted future scenarios in order to avoid comparability issues in prospective LCA studies (harmonization). As shown by the work of Mendoza Beltran et al. (2018), substantial parts of the underlying LCI database may be modified when generating scenario LCI databases. This raises the question as to how scenario databases and their generation process shall be documented to ensure transparency and reproducibility, e.g. satisfying the FAIR (findable, accessible, interoperable, reusable) principles (Wilkinson et al. 2016) (a suggestion has already been made in the FUTURA framework (Joyce and Björklund 2021)).
– *Access to scenario databases:* Solutions will have to be found to make scenario LCI databases practically available to LCA practitioners, for which, amongst others, data ownership and licenses need to be considered.
– *Guidance for LCA practitioners:* While it is unlikely that the LCA practitioner has full knowledge of all underlying models that led to the generation of a scenario LCI database, he/she may be held responsible for any conclusions derived from the use of the latter in LCA studies. For this reason, it is important that LCA practitioners are guided concerning, amongst others, the following questions: how can LCA practitioners understand what scenario databases represent and which scenarios to use in a specific situation? How can LCA practitioners understand, which changes have been made to the underlying LCI database? When is it meaningful to use uncertainty information derived for current LCI databases, e.g. pedigree values, in scenario databases? How shall the use of scenario databases in prospective LCA studies be reported?
– *Support by LCA software:* As shown also within this paper, LCA software is a crucial factor for enabling the use of scenarios in LCA. The key question is thus: how can LCA software support practitioners most effectively in prospective LCA, including the use of future scenarios? Some ideas for this have been implemented in the Activity Browser, but surely further improvements could be made to satisfy the needs of LCA practitioners.

These and likely other challenges should be addressed jointly by scenario generators, data owners, and LCA software providers considering the practical needs of LCA practitioners.

5 Conclusions

Life cycle inventory databases that represent future scenarios based on a combination of data from existing LCI databases and various scenario sources such as integrated assessment models have recently been developed. Although it is impossible to make precise predictions of the future, these scenario databases fill an important gap for prospective LCA by providing temporally consistent background data when assessing technologies at a future point in time.

This paper presents the superstructure approach, which is a solution to the modelling problems that arise from having a number of background LCI databases (one for each scenario and reference year) instead of just one. The solution consists of converting the individual databases into a superstructure database and an associated scenario difference file, which together can be used to represent different future scenarios. The advantage of this approach is that LCA practitioners can use a single background database and do not have to re-link their foreground systems to different background databases. The approach also facilitates fast and automated LCA calculations for all scenarios and even combinations of foreground and background scenarios. Finally, it also reduces the required disk space.

The paper presents an implementation of the approach in the Activity Browser open source LCA software, which builds on top of the brightway LCA framework and thereby provides not only the proof of concept for the

approach, but a practical tool that anyone in the LCA community can directly apply. The authors are happy to share the superstructure and scenario difference file generated within this paper on demand.

While the presented work represents a technical milestone, further challenges need to be overcome to make the use of future background scenarios more widespread in the LCA community. Solutions are required to enable regular LCA practitioners to access and practically use future scenarios as well as to guide them along the way.

We expect that more future scenarios will be developed by the LCA community, including scenarios for specific sectors and regions, and we hope that these scenarios will be used in prospective LCA studies to make them more meaningful and to make a real difference for guiding our future technology landscape towards sustainability.

References

Arvidsson R, Tillman A-M, Sandén BA, Janssen M, Nordelöf A, Kushnir D, Molander S (2018) Environmental assessment of emerging technologies: recommendations for prospective LCA. J Ind Ecol 22:1286–1294. https://doi.org/10.1111/jiec.12690

Cox B, Bauer C, Mendoza Beltran A, van Vuuren DP, Mutel CL (2020) Life cycle environmental and cost comparison of current and future passenger cars under different energy scenarios. Appl Energy 269:115021. https://doi.org/10.1016/j.apenergy.2020.115021

Cucurachi S, van der Giesen C, Guinée J (2018) Ex-ante LCA of emerging technologies. Procedia CIRP 69:463–468. https://doi.org/10.1016/j.procir.2017.11.005

Gibon T, Wood R, Arvesen A, Bergesen JD, Suh S, Hertwich EG (2015) A methodology for integrated, multiregional life cycle assessment scenarios under large-scale technological change. Environ Sci Technol 49:11218–11226. https://doi.org/10.1021/acs.est.5b01558

Harpprecht C, Oers Lv, Northey S, Yang Y, Steubing B (2021) Environmental impacts of key metals' supply and low-carbon technologies are likely to decrease in the future. J Ind Ecol https://doi.org/10.1111/jiec.13181

Heijungs R (2015) Topological network theory and its application to LCA and IOA and related industrial ecology tools: principles and promise. J Environ Account Manag 3:151–167. https://doi.org/10.5890/JEAM.2015.06.005

Heijungs R, Suh S (2002) The computational structure of life cycle assessment. Eco-efficiency in industry and science. Kluwer, Dordrecht, The Netherlands

Hellweg S, Canals LMI (2014) Emerging approaches, challenges and opportunities in life cycle assessment. Science 344:1109–1113. https://doi.org/10.1126/science.1248361

Hertwich EG et al (2015) Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. PNAS 112:6277–6282. https://doi.org/10.1073/pnas.1312753111

IEA (2020) Energy Technology Perspectives 2020 Paris

IPCC (2013) Climate Change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1017/CBO9781107415324

IPCC (2014) Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland

ISO (2006) ISO 14040. Environmental management — life cycle assessment — principles and framework vol ISO 14040:2006(E). International Standardization Organization, Geneva, Switzerland

Joyce PJ, Björklund A (2021) Futura: a new tool for transparent and shareable scenario analysis in prospective life cycle assessment. J Ind Ecol. https://doi.org/10.1111/jiec.13115

Lesage P, Mutel C, Schenker U, Margni M (2018) Uncertainty analysis in LCA using precalculated aggregated datasets. Int J Life Cycle Assess 23:2248–2265. https://doi.org/10.1007/s11367-018-1444-x

Meide M, Yang Y, Steubing B (submitted) Environmental impacts of future cobalt supply scenarios. J Ind Ecol

Mendoza Beltran A et al (2018) When the background matters: using scenarios from integrated assessment models in prospective life cycle assessment. J Ind Ecol. https://doi.org/10.1111/jiec.12825

Moss RH et al (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747–756. https://doi.org/10.1038/nature08823

Mutel C (2017) Brightway: an open source framework for life cycle assessment. JOSS 2(12), 236. https://doi.org/10.21105/joss.00236

Mutel C (2020) Wurst. GitHub. https://github.com/polca/wurst. Accessed 07 Dec 2020

NEEDS (2009) New Energy Externalities Developments for Sustainability Project. The European reference life cycle inventory database of future electricity supply systems. http://www.needs-project.org/needswebdb/. Accessed Dec 2020

O'Neill BC et al (2014) A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim Change 122:387–400. https://doi.org/10.1007/s10584-013-0905-2

Pizzol M, Sacchi R, Köhler S, Anderson Erjavec A (2021) Non-linearity in the life cycle assessment of scalable and emerging technologies. Front Sustain 1:13. https://doi.org/10.3389/frsus.2020.611593

Rockström J et al (2009) A safe operating space for humanity. Nature 461:472–475

Sacchi R, Bauer C, Cox BL (2021) Does size matter? The influence of size, load factor, range autonomy, and application type on the life cycle assessment of current and future medium- and heavy-duty vehicles. Environ Sci Technol. https://doi.org/10.1021/acs.est.0c07773

Sacchi R et al (submitted) Prospective Environmental Impact assessment (premise): a streamlined approach to producing databases for prospective life cycle assessment using integrated assessment models. Renew Sust Energ Rev

Stehfest E, van Vuuren D, Kram T, Bouwman L, Alkemade R, Bakkenes M, Biemans H, Bouwman A, den Elzen M, Janse J, Lucas P, van Minnen J, Müller C, Gerdien Prins A (2014) Integrated assessment of global environmental change with IMAGE 3.0: Model description and policy applications. The Hague: PBL Netherlands Environmental Assessment Agency

Steubing B, de Koning D, Haas A, Mutel CL (2020) Softw Impacts 3:100012. https://doi.org/10.1016/j.simpa.2019.100012

Thonemann N, Schulte A, Maga D (2020) How to conduct prospective life cycle assessment for emerging technologies? A systematic review and methodological guidance. Sustainability 12:1192

Tsoy N, Steubing B, van der Giesen C, Guinée J (2020) Upscaling methods used in ex ante life cycle assessment of emerging technologies: a review. Int J Life Cycle Assess. https://doi.org/10.1007/s11367-020-01796-8

van der Giesen C, Cucurachi S, Guinée J, Kramer GJ, Tukker A (2020) A critical view on the current application of LCA for new technologies and recommendations for improved practice. J Clean Prod 259:120904. https://doi.org/10.1016/j.jclepro.2020.120904

van Vuuren DP et al (2014) A new scenario framework for Climate Change Research: scenario matrix architecture. Clim Change 122:373–386. https://doi.org/10.1007/s10584-013-0906-1

Vandepaer L, Gibon T (2018) The integration of energy scenarios into LCA: LCM2017 Conference Workshop, Luxembourg, September 5, 2017. Int J Life Cycle Assess 23:970–977. https://doi.org/10.1007/s11367-017-1435-3

Villares M, Işıldar A, van der Giesen C, Guinée J (2017) Does ex ante application enhance the usefulness of LCA? A case study on an emerging technology for metal recovery from e-waste. Int J Life Cycle Assess 22:1618–1633. https://doi.org/10.1007/s11367-017-1270-6

Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B (2016) The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 21(9):1218–1230. https://doi.org/10.1007/s11367-016-1087-8

Wilkinson MD et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.