A model for streamlining and automating Path Exchange hybrid life cycle assessment

André Stephan, Robert H. Crawford[1], Paul-Antoine Bontinck

Faculty of Architecture, Building and Planning, The University of Melbourne, Victoria 3010, Australia

[1]Corresponding Author

Tel: +61383448745

e-mail: rhcr@unimelb.edu.au

ORCID: 0000-0002-0189-3221

Abstract

Purpose

Life cycle Assessment (LCA) is inherently complex and time consuming. The compilation of life cycle inventories (LCI) using a traditional process analysis typically involves the collection of data for dozens to hundreds of individual processes. More comprehensive LCI methods, such as input-output analysis and hybrid analysis can include data for billions of individual transactions, or transactions/processes, respectively. While these two methods are known to provide a much more comprehensive overview of a product's supply chain and related environmental flows, they further compound the complex and time-consuming nature of an LCA. This has limited the uptake of more comprehensive LCI methods, potentially leading to ill-informed environmental decision-making. A more accessible approach for compiling a hybrid LCI is needed to facilitate its wider use.

Methods

This study develops a model for streamlining a hybrid LCI by automating various components of the approach. The model is based on the Path Exchange hybrid analysis method and includes a series of inter-related modules developed using object-oriented programming in Python. Individual modules have been developed for each task involved in compiling a hybrid LCI, including data processing, Structural Path Analysis, and path exchange or hybridisation.

Results and discussion

The production of plasterboard is used as a case study to demonstrate the application of the automated hybrid model. Australian process and input-output data are used to determine a hybrid embodied greenhouse gas emissions value.

Full automation of the *node* correspondence process, where *nodes* relating to identical processes across process and input-output data are identified, remains a challenge. This is due to varied dataset coverage, different levels of disaggregation between data sources, and lack of detail of activities and coverage for specific processes. However, by automating other aspects of the compilation of a hybrid LCI, the comprehensive supply chain coverage afforded by hybrid analysis is able to be made more accessible to the broader LCA community.

Conclusions

This study shows that it is possible to automate various aspects of a hybrid LCI in order to address traditional barriers to its uptake. The object-oriented approach used enables the data or other aspects of the model to be easily updated to contextualise an analysis in order to calculate hybrid values for any environmental flow for any variety of products in any region of the world. This will improve environmental decision-making, critical for addressing the pressing global environmental issues of our time.

Keywords: Path Exchange hybrid; automation; life cycle assessment; life cycle inventory.

1. Introduction

Our global economy is applying ever-increasing pressure on the Earth's natural resources and ecosystems. According to data from the Global Material Flows Database (Vienna University of Economics and Business 2017), 84 billion tonnes of minerals, ores, fuels and biomass were extracted to sustain our needs in 2013. To put things in perspective, this represents over 11 tonnes of material extracted per capita on an annual basis (World Bank Group 2017). Minerals used for the construction sector alone represent 44% of this total (Vienna University of Economics and Business 2017). Even without considering the fuel, biomass and metal ore requirements of the construction sector, this means that construction the single most material-intensive sector of our global economy. The extraction, transport, processing, assembly, replacement and decommissioning of construction materials results in significant environmental flows, such as energy use, greenhouse gas emissions, and water use. These so-called 'embodied flows' induce a broad range of environmental effects such as resource depletion, air pollution and climate change. However, embodied flows are often significantly underestimated.

Life cycle assessment (LCA) can be used to quantify these embodied environmental flows and effects. A key element of an LCA, the compilation of a life cycle inventory, which sums all flows coming in and out of the system under study, can be undertaken using one of three methods: process, input-output or hybrid analysis. Each of these methods has been described in detail in the literature (inter alia Crawford et al. 2018a; Lenzen 2000; Suh et al. 2004), and a detailed review of their strengths and weaknesses falls outside of the scope of this work. Process analysis is usually referred to as a bottom-up approach, whereby specific data is collected for specific processes. As such, it suffers from a systematic truncation of the system boundaries (Majeau-Bettez et al. 2011), as it is almost impossible to collect specific data representative of all aspects of any given system. Input-output analysis, which is referred to as a top-down approach, uses data collected from statistical agencies and typically represents every transaction taking place between every sector of a national economy (Leontief 1970). This high level approach also means that the granularity of the data is generally coarse (Lenzen 2000). Information is generally provided at the sector level, and more detailed analyses are often difficult, such as a comparison of practices within a specific sector. Hybrid analysis uses both process and input-output data in a way that reinforces their individual strengths – specificity and completeness, while aiming to mitigate their respective weaknesses (Suh et al. 2004). However, this comes at a price of increased complexity and time requirements. The benefits associated with a more complete system boundary in hybrid analysis was recently questioned (Yang 2017a; Yang et al. 2017). This resulted in a discussion between hybrid life cycle assessment scholars (Gibon and Schaubroeck 2017; Schaubroeck and Gibon 2017; Yang 2017b). The discussion determined that while a more complete system boundary resulting from hybrid analysis might not improve the accuracy of the results from a theoretical perspective, it does provide a more comprehensive understanding of the environmental effects associated with a product or process. Another important outcome was the significance of the data used, not only the life cycle inventory approach. Subsequently, Pomponi and Lenzen (2018) demonstrated that when using more realistic process and input-output data (as opposed to theoretical data with limited representativeness of real systems in Yang (2017a)), hybrid analysis is likely to improve both accuracy and system completeness. It is important to note that the more realistic data is obtained by modifying the technological process matrix to introduce more interdependencies as in a typical process database (e.g. ecoinvent). Similarly, the input-output A matrix is modified to increase its eigenvalue, in line with the actual eigenvalues of A matrices for real economies. Testing hybrid analysis on actual process and input-output data would provide a more robust demonstration of its capacity to increase accuracy and system completeness. Nevertheless, and for the reasons outlined above, hybrid analysis, although more complex and data-intensive, tends to provide additional benefits over both process and input-output analysis while reducing some of the limitations associated with each individual approach.

Current research in the building industry suggests that the use of process analysis can introduce a significant bias to the conclusions of a LCA (Crawford 2011). With energy use being one of the most significant environmental flows associated with buildings (Junnila 2004; Oregi et al. 2015), comparing the embodied energy associated with increased operational energy efficiency becomes critical information for decision-making. Studies using hybrid analysis, such as Crawford et al. (2016), show that current building energy efficiency regulations could paradoxically result in increased life cycle energy use because the additional embodied energy is not necessarily offset by operational energy savings. This contradicts the results often found when using the less comprehensive process analysis method to quantify embodied energy (see for example Tettey et al. (2014)). In addition, with the push for greater operational energy efficiency, as well as the advent of net-zero operational energy buildings, operational energy is declining, relatively increasing the contribution of embodied energy (Blengini and Di Carlo 2010). Therefore, comprehensive quantification of embodied energy, using broad system boundaries, is becoming increasingly important, especially when it comes to buildings.

Hybrid analysis provides the most comprehensive quantification of embodied flows (Dixit 2017). Crawford (2011) and Stephan and Stephan (2014) have shown that hybrid analysis can result in embodied energy figures up to four times higher than when using process analysis. However, hybrid analysis, notably the Path Exchange method for hybridisation (see Section 2.1), has typically been applied in academic settings, failing to make its way to more mainstream applications of life cycle assessment. The complexity of this method and the absence of computational tools for its application mean that only a small group of researchers and practitioners possess the means to apply it. By overcoming these barriers, hybrid analysis would be more accessible and more likely to inform environmental decision-making, particularly in the construction industry. The work described here focuses on the development of an automated approach to the application of the Path Exchange hybrid analysis method, as first theorised by Treloar (1997), and formalised by Lenzen and Crawford (2009).

The Path Exchange hybrid analysis method is chosen because it naturally streamlines the hybridisation process compared to a full matrix integration and avoids the need for a number of assumptions to be made (see Section 2.1).

1.1. Aim and scope

The aim of this paper is to describe the development of an automated model for the application of the Path Exchange hybrid analysis method**.** This model will help overcome some of the barriers limiting the broad application of the Path Exchange hybrid analysis method for compiling a LCI. This will allow researchers and practitioners to compile hybrid coefficients of embodied flows for a large range of basic products, such as construction materials, and compile detailed LCIs of more complex products. The paper discusses the development of this model in the context of calculating a hybrid value for a basic product, using plasterboard as a case study. This paper extends on the previous work of Crawford et al. (2017), which discusses the need for automating the Path Exchange hybrid analysis method and proposes the framework that underpins the model.

This paper is structured into six sections. Section 2 introduces the Path Exchange hybrid analysis method that is used within the model, as well as the general framework of the model that is developed. Section 3 describes the model itself and guides the reader through the different steps of a hybrid analysis and how these have been implemented within the model. Section 4 demonstrates the application of the model using plasterboard as a case study. Section 5 discusses the potential for broader application of the model, current limitations, and areas for further improvement, before concluding in Section 6.

2. Method

This section provides an overview of the Path Exchange hybrid analysis method, before describing the development of the model for automating the compilation of a hybrid LCI using this approach.

2.1. Path Exchange hybrid analysis method

The Path Exchange (PXC) hybrid analysis method has been extensively described in the literature (Lenzen and Crawford 2009; Treloar 1997; Treloar 1998). This paper does not review its theoretical underpinnings, but provides a step-by-step outline of the approach, depicted in Figure 1, which sets the context for the model detailed in Section 3.

The PXC method is based on the assumption that process and environmentally-extended input-output models are, generally speaking, two different approaches to describe similar supply chains. While this is true in principle, differences in the base models and resolution mean that a one-to-one equivalence is often difficult to establish between data points of the two models.

Prior to compiling a PXC-based LCI, data must be collected and formatted appropriately, as shown in Step 1 and 2 of Figure 1. This includes the compilation of process and input-output direct coefficient matrices for process and input-output inventories, and the calculation of total intensities for all processes and sectors represented in these matrices. This is referred to as data processing.

The first step of the PXC method involves conducting a Structural Path Analysis (SPA) on the process and corresponding input-output sector analysed (Figure 1, Step 3). The SPA was first described in economics in 1984 (Crama et al. 1984; Defourny and Thorbecke 1984), and can be used to analyse all contributions to a particular output from the entire supply chain. The output of a SPA is a series of mutually exclusive *nodes*, each representing a good or service provided from one tier to another within the supply chain analysed. A series of *nodes*, corresponding to a chain of *transactions* leading to the sector being assessed, is referred to as a *pathway* – for instance iron ore purchased by the reinforcing steel bar sector, in turn purchased by the residential construction sector for an apartment building. Each *node* contains data on its associated environmental flows based on either physical or financial quantities (e.g. GJ of energy per kg of material or GJ of energy per currency unit of material). Based on the equivalence assumption stated above, it is possible to highlight corresponding relationships between processes and sectors in the two models. This enables the substitution of top-down input-output data with generally more precise, bottom-up process data.

The SPA is used to replace average sector-wide input-output *nodes* with specific process *nodes* (and the associated environmental flow data), thus making the model more specific to the product being analysed while maintaining the system boundary completeness provided by the input-output model. To replace these *nodes*, the process and input-output supply chain unveiled by the SPA must be analysed in detail to identify potential correspondences between their *pathways* (Figure 1, Step 4).

<Take in Figure 1>

Once this process of finding correspondences has been completed, the hybrid value for specific environmental flows can be calculated. The direct intensities associated with all identified process *nodes* are summed, representing the process component of the hybrid value (Figure 1, Step 5). Likewise, the direct intensities of the input-output *nodes* for which process data is available are summed, and then subtracted from the total intensity of the relevant input-output sector. The output of this calculation is converted from an intensity per financial output (e.g. GJ/$) to an intensity per physical output (e.g. GJ/tonne), based on the economic value (i.e. cost or base price without profit margins) of the product modelled (Figure 1, Step 6). It is referred to as the remainder – the input-output component of the hybrid value. The process and input-output components are summed to obtain the hybrid value (Figure 1, Step 7).

The main advantage of the PXC hybrid analysis method is that it operates on mutually exclusive process and input-output *nodes* and that it replaces only the most significant *nodes*. Unlike the integrated hybrid method which combines the entire process and input-output matrices into on larger matrix (and accounts for loops in supply chains), the changes to a supply chain are discrete, at the node level, and do not require other changes in the matrix. This avoids the need to collect data and make assumptions that would be needed to populate the so-called upstream and downstream cut-off matrices in an integrated hybrid model (Baboulet 2009; Suh and Huppes 2000). That makes the process more efficient as only the most significant *nodes* are modified. These are the reasons behind the choice to automate the PXC hybrid analysis method. For a detailed review of hybrid analysis methods, refer to Crawford et al. (2018a).

2.2. Automating Path Exchange hybrid analysis

The PXC method, as described in Section 2.1 and Figure 1, involves steps which rely on mathematical computation. Many of these steps are currently performed manually, which requires specialist expertise and has thus limited the widespread use of this approach. This represents an ideal opportunity for automation using computer programming in order to streamline PXC hybrid analysis to encourage its broader uptake.

The computer language selected for the development of the automated PXC model is Python (Python Software Foundation 2017), a general purpose language which enables a wide range of tasks through its use of libraries. Powerful libraries can be imported to facilitate the analysis. For instance, *pandas* (Augspurger et al. 2017)*,* which is a data analysis library that provides a multitude of data structures that improve and facilitate matrix calculations. Other relevant libraries include *wxPython* (2018), which assists with designing and producing graphical user interfaces. Python is an open-source language. Once published, the code can be used or extended by anyone. It works by writing object-oriented code (Lutz 2013), which defines objects – a data structure containing variables referred to as *attributes* and dedicated functions referred to as *methods*, which can be used to perform specific actions (Alic et al. 2016; Lutz 2013). This helps to develop a flexible and versatile model that can be easily updated, built upon, and reviewed by peers. In turn, this flexibility produces outputs useable by a wider audience and code that is adaptable to a range of purposes. For example, a *SupplyChain* object can be used to conduct a SPA and extract its *pathways* and *nodes*. This *SupplyChain* object can also be used to represent data visually by adding relevant visualisation methods. The model is built using best practice for scientific models described by Dubois (2005), Wilson et al. (2014) and Pauliuk et al. (2015). It also goes beyond these best practice models by implementing additional measures for code readability, such as adhering to the recognised style guide for Python code, i.e. PEP 8 (van Rossum et al. 2018). This ensures the production of code that is of the highest possible standard, and that can be used by the scientific community for a variety of tasks.

Developing code for the hybridisation of process and input-output data using the PXC method involved a number of critical steps. In the first instance, the PXC method is reviewed in detail and divided into a series of high level processes – e.g. data processing, Structural Path Analysis, etc. This forms the basis of the architecture of the model, where one *module* (a Python file containing *classes* and *functions*, defined in point 3 below) is created for each process in the form of a Python file. These include '*preprocessing.py*', '*spa.py*' and '*hybridisation.py*'. The high level architecture of the model is depicted in Figure 2 and a detailed overview of the model is provided in Section 3. Each *module* is developed so that it can be used in conjunction with the others, or as a standalone feature. The development of each *module* follows a similar approach:

1. **Define an expected outcome from the *module***, in terms of file output, data format, objects to model, etc. For instance, the output of the data processing *module* was defined as an *'.lci'* file object which includes, amongst other objects and variables, a direct coefficient matrix, stress matrix, and impact assessment method.
2. **Detail individual tasks required to complete a *module* and reach the expected outcome**. These can be in the form of a flow diagram, or a list of actions. In the case of the data processing *module*, the inputs include data collected either from software such as Simapro (PRé

Sustainability 2018), from life cycle inventory database providers (e.g. *ecoinvent* (Wernet et al. 2016)), or as a supply-use table collected from the IELab (Lenzen et al. 2014). The individual tasks required to process the IELab output for example, include the conversion of the supply-use table into an input-output table, the calculation of the Leontief inverse, and the calculation of the vectors of direct requirements. Similarly, a list of tasks is drawn to process all other forms of data used by the model.

3. **Design a set of *classes* and their *methods* as well as *functions* in order to complete each individual task**. Taking the example of the Structural Path Analysis (SPA) *module*, a series of nested *classes* (a class is a representation of an object in code) are defined to assess the different aspects of a supply chain, including *nodes*, *transactions*, and *pathways*. Examples of *methods* include '*extract_pathways*', which extracts all *pathways* in a list, or *'export_to_csv'*, which is used to convert the *pathways* extracted into a csv file. A *function* is a general process that receives inputs, follows a series of processes and often generates an output. An example in the code is the '*matrix_to_csv' function*, which receives a matrix object and file path, and returns a csv file in which the matrix has been saved.
4. **Test and debug the *module* and its outcome**. As issues arise during the testing phase, *modules* are modified, new *functions* are added and new *methods* are implemented in *classes* to deal with each issue. Thus *modules* as a whole evolve towards their final structure. Debugging may involve running the model on dummy data and manually recreating the calculation to ascertain that an appropriate result is reached, or running tests with data that has previously been used and where the result is known. This stage is where the iterative aspect of programming is the most prominent, as the role of *modules*, *classes* and *functions* are modified over time.

<Take in Figure 2>

Two additional *modules* were created and linked to this high level structure, as depicted in Figure 2. They are used as toolboxes throughout the entire model. The first one, *'generic.py'* (1), houses *functions* that are used in different parts of the model, facilitating the reuse of existing code rather than writing multiple versions of it, in accordance with current programming best practice. The second, *'config.py'* (2), contains all hard-coded variables used in all other *modules*, thus providing a central repository for version control.

With computer programming being an iterative process by definition, a strict version control system was implemented, as recommended by Wilson et al. (2014). Among existing tools, *SourceTree* was selected (Atlassian 2017), as it allows parallel development of code by several programmers, while keeping track of all changes made to the code and Master version on a cloud-based platform. This is particularly helpful in the code testing phase, where bugs inadvertently introduced to later versions of the code can be identified. Individual and independent modules are developed, tested and corrected without interfering with the rest of the model. Over time, the architecture of the model is reviewed and updated, as the potential for simplification and clarification is identified by the team.

Since aspects of the PXC hybrid method cannot be entirely automated, as discussed previously (Bontinck et al. 2017; Crawford et al. 2017), a user interface was developed in parallel. This provides an opportunity for users to enrich the outcome of the model with their specialist knowledge, and also facilitates the use of the model by users who may not be Python specialists. In addition, having a user-interface also enables a standardisation of the way the model is used, making its output more robust. It also provides an opportunity to compile the model as a standalone executable file later on. This makes it portable and useable by anyone as it does not rely on library dependencies.

3. Development of the automated Path Exchange hybrid model

This section provides a step-by-step description of the model developed for the automation of the PXC hybrid method. A detailed video walkthrough is available in open-access on FigShare (https://doi.org/10.6084/m9.figshare.5687773, (Stephan 2018)).

3.1. Collection of input-output and process data

The model is capable of using any source of process or input-output data, as long as these have been processed accordingly. In terms of process data, any library can be extracted from LCA software, such as Simapro (PRé Sustainability 2018), as a technology matrix and used within the model. Alternatively, a database of *ecospold* files can be used, as provided by *ecoinvent* (Wernet et al. 2016). In this case, the database must first be processed using the *ecospold2matrix* module developed by Majeau-Bettez (2014), a Python model that can be used to consolidate a database of *ecospold* files into a set of matrices. Any other technological matrix can be used, as long as it has been compiled into a matrix of direct coefficients. In terms of input-output data, the model can use either supply-use tables which is the typical output of the input-output models in the Australian Industrial Ecology Virtual Laboratory

(IELab) (Lenzen et al. 2014), or symmetric input-output tables produced from these supply-use tables. The IELab provides highly disaggregated input-output data in the Australian context, with up to 1,284 individual sectors described within the matrix.

3.2. Processing input-output and process data

The aim of the data processing *module* is to prepare the information collected for the Structural Path Analysis (SPA) *module*. A simplified overview of this process is depicted in Figure 3.

<Take in Figure 3>

It is worth noting that the *functions* used to process input-output and process data vary. In particular, the calculation of the matrix of direct coefficients and vectors of direct requirements vary depending on whether process or input-output data is being processed.

The model uses symmetric input-output tables to conduct an SPA. It converts supply-use tables, as collected from the IELab to symmetric input-output tables (see Eurostat (2008) and Suh et al. (2010) for detailed information and equations). Supply-use tables, as defined by Eurostat (2008), show the relationship between industries and commodities by reporting on the supply of commodities by industries to the economy, and the use of commodities by industries to produce a certain output. Although this system provides valuable information, symmetric input-output tables were preferred because the additional information does not serve a clear purpose when it comes to conducting a PXC hybrid analysis and can provide unnecessarily complex results. Notwithstanding, the SPA code developed also works with supply-use tables should these be used to conduct a PXC hybrid analysis in the future.

In the case of a technology matrix extracted from Simapro, a normalisation step based on the output represented by each process is enough to obtain a matrix of direct coefficients. The information required for this calculation is included within the technology matrix when extracted from Simapro. If using a database of *ecospold* files, the matrix of direct coefficients is obtained from the *ecospold2matrix* module (Majeau-Bettez 2014) and no further calculations are required.

The calculation of direct requirements vectors varies slightly as well. The 'satellite' vectors typically used in input-output analysis generally represent fully characterised indicators (such as greenhouse gas (GHG) emissions, in kg CO_2e). On the other hand, the 'stress' matrix obtained from a process model includes all individual flows – in the case of GHG emissions this would include carbon dioxide, methane, dinitrogen monoxide, etc. The stress matrix is converted to a series of characterised indicator vectors using an impact assessment matrix. In the case of GHG emissions, this matrix will provide an equivalence between raw flows (*e.g.* in kg CH_4) and characterised flows (*e.g.* in kg CO_2e). All other steps follow the same principle for all types of data, from that point onwards.

It is important to highlight the need to use process and input-output data that are representative of the same period. This is to improve the equivalence between the process and input-output supply chains. While the model will accept any base input-output and process data, we encourage the use of datasets that have at most five years of difference for when data was collected. In the best scenario, both datasets will be representative of the same year. However, even if that is the case, it is almost impossible in practice to ensure that a particular process pathway is replacing an input-output pathway of the same period. This is because within a process or input-output database, not all processes or transaction data are for the year of release, but are instead collected over many years and revised on an ongoing basis. There are therefore temporal discrepancies within process and input-output databases, in addition to between process and input-output databases. Unless every process and transaction has accompanying metadata that provides information about its temporal representativeness, it will be virtually impossible to ensure that the Path Exchange technique is swapping process and input-output data that are for the same year. Given the current quality of data, aiming for a minimal difference in temporal representativeness for the two datasets is good practice.

3.3. Defining a product and selecting databases (Figure 1, Steps 1-2)

Once datasets have been processed, a user can define a product in the graphical user interface (see Figure 4, 'Define Product' box and Figure 1, Step 1). This consists of providing a name (e.g. plasterboard), category (e.g. minerals), type (e.g. plaster), a country of production (e.g. Australia) and a functional unit (e.g. m^2). In addition, the user must specify the cost and weight of the product per functional unit. For some products, such as building materials, a user can specify specific heat, density and thermal conductivity characteristics. For each of these parameters, the user must indicate the year and source of the data. This significantly improves transparency and aligns with good data management practices. Once these parameters are defined, the user can select a process database and an input-output database (Figure 1, Step 2) in order to conduct the Structural Path Analysis.

<Take in Figure 4>

3.4. Conducting the Structural Path Analysis (Figure 1, Step 3)

After defining the product, selecting the process and input-output databases and loading them, the user needs to select a relevant process and sector from the relevant list. The process and sector selected need to reflect as much as possible the defined product. Usually, the exact process of production can be identified and with higher sectorial disaggregation of input-output data, a closely matching sector can also be selected (see Figure 4, 'Select process' and 'Select sector' boxes). Once the relevant process and sector are selected, the user specifies the SPA parameters (i.e. the environmental flows to consider, and the cut-off threshold and maximum stage upstream for each of the process and input-output data), before conducting the SPA (Figure 1, Step 3).

The Structural Path Analysis *module* (*spa.py*) is the heart of the model and a core step in the PXC method. The code was inspired by the algorithm developed by the late A/Prof. Graham Treloar during his PhD (Treloar 1998), but was completely revised.

The main class is the *SupplyChain* class. It receives the technological matrix from the relevant process or input-output life cycle inventory file (.lci), as well as direct and total environmental intensity vectors (multipliers in input-output terms) for each flow, e.g. energy, water, greenhouse gas emissions, pollutants, etc. The specified number of stages to explore in the supply chain and the cut-off threshold values to prune pathways are also used by this class.

The *method* '*extract_pathways'* is called to generate *pathway* objects, which in turn, are constituted of lists of *nodes* and *transactions*. This nested model is a more robust reflection of reality (compared to simply extracting values) and enables the implementation of a myriad of functionalities within the code. The *pathway* extraction proceeds as follows:

1. The code reads the maximum stage value and dynamically generates *methods* for the *SupplyChain* class to extract *pathways* at the nth stage. This is a major innovation of the code, which enables the user to extract as many stages upstream as they wish. It is particularly useful for detailed process models (e.g. *ecoinvent*) where environmental flows can often occur far upstream in the supply chain.
2. The code iterates through the technological inputs from other processes/sectors into the target process/sector (through the rows of the target process/sector's column in the technological matrix).

3. The technological coefficient of an input process/sector is multiplied by the associated total environmental intensity, for each flow.
4. If the intensity is larger than, or equal to the imposed threshold for that flow, the inputs into that sector are considered, as long as the stage of the supply chain is less than, or equal to the specified maximum stage.
5. When either the threshold or the maximum stage condition is not satisfied, the rest of the *pathways* are pruned for that specific flow, meaning that no intensities are saved in upstream *node* objects for that flow.
6. When no further flow satisfies the threshold condition or the maximum stage has been reached, the *pathway* object is saved in the list of pathways of the *SupplyChain* object.

It is important to note that by covering all flows in a single pathway extraction, the algorithm saves time compared to extracting pathways for each flow, individually. This is because many *nodes* are relevant for more than one flow, notably for the first few stages upstream. Capturing multiple flows in one pathway extraction also significantly streamlines adding new environmental satellites to the model.

This so-called ‘depth first search’ is widely used in applications involving tree structures (such as supply chains), namely to save memory, when the entire search space is too vast. In this case, using the flow thresholds and stage limit is necessary to limit the number of pathways returned as these would be unnecessarily complex otherwise, and progressively insignificant.

The user can easily visualise what proportion of the total environmental intensity of a flow for the target process/sector is covered by the Structural Path Analysis and tweak the input variables in order to obtain a suitable coverage (we encourage the highest coverage available, where possible). Figure 4 depicts the user interface that is used to conduct the Structural Path Analysis on both the target process and sector, simultaneously.

3.5. Identification of corresponding *nodes* (Figure 1, Step 4)

The identification of corresponding input-output and process *nodes* is based on three assumptions.

1. an input-output and a process *node* are equivalent, if representing the same process at the same point in the supply chain;
2. it is preferable to use process data when available, as it is typically more specific to the product analysed; and

3. differences in process and input-output data structures mean that the relationship between *nodes* could be one-to-one, many-to-one, one-to-many or many-to-many.

The model provides support to the user with the use of a concordance dictionary between the list of processes in the technological matrix, and the list of input-output sectors for the classification used. Note that dictionaries are programming data-structures that map a single key to a value. The key can be a number (e.g. the number of an input-output sector in a matrix) or a string (e.g. 'plasterboard'). The value can be any data-structure (e.g. a list or another dictionary) or other object (e.g. a *Pathway* object). When selecting a process *node* in the graphical user interface, the model can identify and list all possible corresponding input-output *node* candidates, and vice-versa. This concordance is being augmented by flow-specific concordance models to facilitate matching environmental flows further upstream in the process model with environmental flows that are typically visible at Stages 2-3 of an input-output SPA. The model as it currently stands still relies on the expert knowledge of the user, and on review systems that take place outside the model. However, the object-oriented approach to this model means that a more sophisticated approach can be implemented in the future, as discussed in Section 4.

When identifying matching *nodes*, the user ranks process or input-output *nodes* by contribution to the total requirement for a particular flow. This ensures that the user can easily identify the most significant process and input-output *nodes* for a particular flow and to match them. This also helps ensure that no significant *nodes* are omitted from the SPA. Should the user identify process *nodes* for which no equivalent input-output data is available, or vice versa, this may prompt them to revise the threshold and maximum stage values to cover a higher proportion of the supply chain.

3.6. Hybridisation (Figure 1, Steps 5-7)

Both the identification of corresponding *nodes* and the hybridisation are conducted within the same *module*. While identifying *nodes* is essentially based on human input, the calculations behind the hybridisation are entirely automated.

<Take in Figure 5>

The use of object-oriented coding is an asset here, as it enables the model to build upon the output of previous *modules* and their specificities. *SupplyChain* objects created by the SPA *module* include lists of *pathways* – which themselves include lists of *nodes* and *transactions* as discussed in Section 3.4. The graphical user interface manipulates these lists to create a dictionary of matches. A schematic overview of the hybridisation *module* is provided in Figure 5. The main output is the creation of a *MatchesDictionary* object, which includes the ID reference of the exchange and the list of input-output and process *nodes* that it covers. Using the information in the *SupplyChain* object, and the dictionary created using the hybridisation *module*, the calculation of the hybrid value is relatively simple.

The *MatchesDictionary* streamlines the inclusion of additional environmental flows at a later stage as the hybridisation and the Path Exchange for these flows is automatically calculated. The user might want to identify *nodes* that are significant for that particular flow and to match them, in addition to the *nodes* already matched. This facilitates the addition of multiple environmental flows as data becomes available.

The output of this *module* is a file that includes all relevant lists, as well as all metadata carried over from previous *modules*, and any metadata entered by the user. However, when called from the graphical user interface for hybridisation (Figure 6), the hybridisation *module* does not create a file, but rather a *PathExchange* object with all necessary information that carried forward onto the Interpretation tab of the graphical user interface.

<Take in Figure 6>

3.7. Interpretation

Once the hybridisation is complete, results are displayed in the Interpretation tab (Figure 7). This comprises the hybrid value for each environmental flow studied and a graph summarising the compilation of this value, including the total input-output environmental intensities replaced by process data and excluded process *nodes.* In addition, this tab enables an in-depth analysis of excluded process *nodes* and the input-output remainder (i.e. the fraction of the total input-output environmental intensity that is outside of the threshold of the SPA). This is presented in a dashboard that enables the user to save the results, generate a summary report, display various graphs, and analyse the different environmental flows mentioned above.

<Take in Figure 7>

In addition, the Interpretation tab allows the user to conduct a sensitivity analysis on the value of the product in currency value per functional unit. This critical variable is used to convert the purely financial results of an input-output analysis into physical values. It can have a significant effect on the results and may vary over short periods of time, in different locations, or even between different suppliers. By enabling this sensitivity analysis, the inherent variability in the hybrid value can be integrated seamlessly into the results.

4. Application of the automated Path Exchange hybrid model

In order to demonstrate the application of the model, we use a case study product, namely plasterboard, a common construction material, produced in Australia.

The process database used is the Australian Life Cycle Inventory database (AusLCI) (Grant 2016), which is collected as a technological matrix from Simapro. The input-output database used is collected from the IELab and is representative of the Australian economy in 2014, with a 1,284 sector disaggregation. The analysis focuses on GHG emissions, with a functional unit of kilograms of carbon dioxide equivalent per square metre of plasterboard.

The application of the model to determine the embodied GHG emissions of plasterboard is summarised in Figure 8. The entire analysis, including results of the SPA, as well as the hybridisation, are summarised in an open-access Excel file accessible on FigShare (see Section 5.4).

<Take in Figure 8>

The resulting hybrid greenhouse gas emissions value is 5.33 kg CO_2e/m^2. The process and input-output components of the hybrid value each represent roughly 50% of the final figure. The process *nodes* selected for hybridisation represent 56% of the total pure process intensity, while the input-output *nodes* used as part of the hybrid value represent 53% of the total input-output sector intensity. The pure input-output and process greenhouse gas emissions values are 4.68 kg CO_2e/m^2 and 4.7 kg CO_2e/m^2, respectively.

Overall, 44% of the process model was excluded from the analysis. Of this, processes related to capital infrastructure represent 60%, while transport processes represent 14%. Infrastructure processes were excluded as they are commonly modelled to provide a general representation of capital investment to a process model, but include significant uncertainties about their representativeness. Transport processes were excluded as they are best modelled using proxy values for distances, while input-output data is representative of actual expenditure on transport, and thus might be closer to reality.

The result is directly correlated to the price assigned to the material being analysed. Here, a variation of ± 20% of the cost results in a variation of ± 10% in the hybrid value. As this value will likely drive the result of a hybrid analysis, it is particularly important that the model provides references to and justification for the cost value that is used.

5. Discussion

5.1. Progress made through the implementation of the model

This paper has provided an in-depth description of a model for automating the PXC hybrid method for compiling an LCI. It demonstrates that what is typically a very time intensive process can be significantly streamlined by using advanced object-oriented models that facilitate the extraction of *nodes*, matching of process and input-output *nodes*, interpretation of results, and sensitivity analysis. This is currently demonstrated through anecdotal evidence arising from the authors' experience in manually compiling hybrid LCIs using the PXC hybrid method. Furthermore, this model standardises the Path Exchange process by systematising the compilation of hybrid values. The transparency offered by the model, which conserves all core data and metadata and stores them in relevant virtual objects that replicate actual entities (e.g. a *SupplyChain* object) improves replicability. This helps to overcome one of the most significant challenges found in current life cycle assessments – a lack of transparency in compiling life cycle inventories (Dixit 2017; Säynäjoki et al. 2017).

An important aspect of this model is that it saves hybrid models as individual files comprising all assumptions and decisions made during the analysis. This enables saving different versions of the model in an archive. For instance, it would be possible to archive the versions used for specific analysis for future reference, for review purposes, or for reporting. Each file includes a record of all decisions made throughout the creation of the model and can also include user comments, thus making the analysis entirely transparent, improving reproducibility and enabling more efficient data sharing.

In addition, the automated nature of the model means that entire process databases can be used in the form of technological matrices, instead of using a series of single inventories. This brings the process component of the hybrid model in line with models used by a majority of life cycle assessment practitioners, making comparisons between studies more feasible.

The use of object-oriented programming facilitates carrying over any of the information provided by the model. For instance, the model can analyse the so-called 'input-output remainder', which was previously reported as a single figure covering a variety of material and non-material inputs across the supply chain (e.g. Rauf and Crawford 2015). This is significant as the remainder is typically the largest differentiating factor between process and hybrid results. The breakdown of the remainder into its most significant input-output sectors, sheds light on which elements of the supply chain may warrant further collection of primary process data. Analysing the remainder can also identify potential *nodes* or *pathways* that are irrelevant to the product being analysed and which can be excluded from the study.

In addition, the graphical user interface provides a dynamic overview of the results, as the analysis is being performed. This is an opportunity to implement certain quality requirements, such as a minimum proportion of the process model to be covered in the correspondence assessment, etc. Quality requirements will in turn help produce higher quality and more targeted results due to greater user control, reduce uncertainties and enable more complex analyses using the PXC hybrid method.

For a life cycle assessment practitioner, the developed model could be used as a supplement to existing process analysis software, e.g. Simapro, Open-LCA, etc. as long as this software can export the process system as a matrix. This matrix can then be fed to the data processing module. This would enable a practitioner to first model the process system (including custom process data) on a third party software, before moving to the developed model for hybridisation. Hybridised coefficients are expressed in terms of characterised indicator (e.g. greenhouse gas emissions in kg CO_2e). The model does not currently allow the calculation of final life cycle impact assessment indicators, e.g. health impacts. This would need to be manually conducted by the assessor. However, the model does provide all the relevant process and input-output metadata for that task.

5.2. Situating the model among existing studies

Existing studies that have attempted to streamline hybrid LCIs are used to provide additional context to the model. The SPA model described in this paper departs significantly from existing models, in terms of complexity, functionality and ease of updating. Notably, the SPA model of Treloar (1998) used Visual Basic programming to conduct a SPA in Excel. Treloar's algorithm was implemented to reach a maximum of twelve stages upstream and the model is not object-oriented (despite the capacity to build object-oriented models at that time) and therefore does not allow further automated computations, beyond the extraction of *nodes* and *pathways* and sensitivity analysis. The current hybrid model has the ability to conduct a SPA on any number of stages upstream (dynamically). Together with the object-oriented nature of the SPA *module,* and its support for multi-regional matrices, this makes it one of the most sophisticated SPA models to date.

More recently, there have been attempts to automate hybrid LCIs using the integrated hybrid method (see Crawford et al. (2018a) for a description of the different hybrid methods). While it is mathematically possible to fully integrate a process and input-output matrix into a single larger matrix, updating this model over time can be very time-intensive. Indeed, the upstream and downstream cut-off matrices, used to align requirements between the process and input-output matrices, would need to be recomputed for each product. In comparison, the PXC hybrid method, which focuses on replacing the most significant *nodes* only, is much easier to update. For new datasets using the same process and sector classification as previous versions, the model can simply be rerun while conserving as much of the Path Exchange information as necessary. If new datasets use a different classification, concordance dictionaries can be compiled once and automatically convert the bulk of the Path Exchange information to move between versions.

Finally, this model builds upon the existing efforts of Pauliuk et al. (2015) to streamline the processing of process data. While Pauliuk et al. (2015) convert *ecoinvent ecospold* files into a square technological matrix, we have been able to produce object-oriented architecture, separate algorithms from the data itself, and establish a much broader range of functionalities. This can streamline and enhance the use of existing process datasets, such as those contained in *ecoinvent*.

5.3. Limitations in automating the hybridisation process

There are inherent limitations to restraining a complex analysis such as the PXC hybrid method to the use of an algorithm, as discussed previously (Bontinck et al. 2017; Crawford et al. 2017). There is a fine line between using computing power to assist the analysis and reinforce the robustness of its output, and leaving it entirely up to the algorithm.

Assessing the input-output and process models to identify similarities in their representation of a single supply chain is a complex task in itself and represents a central part of the PXC method (Lenzen and Crawford 2009). As mentioned in Section 1, input-output and process models are built using very different approaches (Suh and Huppes 2005), and a detailed understanding of both approaches is essential in order to identify correspondence between *nodes*. This in itself makes the full automation of the model a complex endeavour. Even if artificial intelligence might be used to fully automate this process in the future (see Section 5.4), this might not be the optimal way forward for the time being.

A more reliable estimation of actual time saved with the use of the model is also required to provide a more precise evaluation of the benefits of automation. This would entail comparing the time required for a statistically significant number of practitioners to manually and semi-automatically compile a hybrid LCI using the PXC method for hybridisation. However, given the limited number of persons that have expertise in the PXC method for hybridisation, testing the time saved from using the model in a reliable manner might not be possible in the short term.

Producing a semi-automated PXC hybrid model has the added advantage of providing the user with an opportunity to learn about supply chains through the analysis, and to have a deeper understanding of the reasons underlying a particular outcome. The systematic nature of LCA means that besides supporting decision-making, it can serve as an educational tool in the sense that it collects and compiles a mass of information in a standardised format that can be understood by non-specialists. Using pre-compiled coefficients without any knowledge of the assumptions used to build them can lead to misunderstandings in the interpretation of results, or a limited understanding of the outcome of an analysis. For instance, the ICE (Inventory of Carbon and Energy) database (Hammond and Jones 2008) compiles embodied energy and GHG emissions coefficients from the literature. It provides pre-compiled coefficients for a range of products (mainly construction materials), and provides some further information such as the spread of coefficients for each product. However, ICE does not provide any information on the LCI methods or assumptions used to compile the coefficients. This means that figures with possibly very different system boundaries, temporal relevance, geographic relevance, and LCI techniques, are averaged. This could lead to a decay of knowledge resulting from a lack of awareness of the limitations of such coefficients. A better solution would be a fully transparent model that contains all the data and metadata used for the compilation of a single hybrid coefficient or value for any product.

In addition, the model does not currently allow the modification of one elementary flow, i.e. the direct requirement for a particular flow associated with one node, e.g. the energy intensity of freight transport per tonne-kilometre. This may be a useful feature to implement in the future.

5.4. Data and code availability

All the data used in the case study are available in an open-access file on Figshare (https://doi.org/10.6084/m9.figshare.5687770, (Bontinck 2018)). In addition, three tables are provided as Supplementary Information, including the ten most significant process and input-output *nodes* (Table S1 and S2), and an example of *node* concordance assessment (Table S3).

The code used to develop the model will be made available on its dedicated Figshare page (https://figshare.com/projects/Improving_the_life_cycle_environmental_performance_of_Australian_construction_projects/27433) (Crawford et al. 2018b).

5.5. Future work

An understanding of the environmental performance of a product across a broad range of environmental flows is critical. This helps to ensure that impacts are not simply shifted as a result of improvements addressing a product's performance relating to a single environmental flow. The model has been developed to enable any number of environmental flows to be analysed, as long as 'satellite' data and a 'stress' matrix is available. While this study has demonstrated the compilation of a hybrid embodied GHG emissions value, the model has been used to produce values of embodied energy and water and work is underway to establish a comprehensive list of coefficients for a wide range of construction materials for these and other environmental flows, such as raw materials, waste and various pollutants.

Identifying corresponding *nodes* within the process and input-output data is the one task that has not yet been fully automated. While manually selected correspondences can be saved for reuse, a more sophisticated suggestion system may be implemented based on a set of general rules for correspondence assessment. This being such a critical aspect of the PXC hybrid method, it would warrant the development of a dedicated module, and some specific guidelines and rules. An example of such a rule might be that input-output *nodes* on logistics are preferred over equivalent process *nodes* as they might be more accurate. Process data on logistics often relies on average values, or on proxy values used to represent transport, but not based on specific primary data. It is expected that the more the model is used, the more rules will be defined.

It is also planned for the model to be capable of analysing more complex products or systems, consisting of multiple products or components, as well as non-material-related processes such as the provision of services. This work is underway, eventually enabling the analysis of the environmental performance of composite products, such as whole buildings or airplanes, extracting detailed physical quantities from project documentation and calculating hybrid embodied environmental flow values by modelling each individual product, or using pre-compiled hybrid coefficients for products (such as construction materials or aeronautic materials), where available.

Another potential avenue is to broaden the model to include the integrated hybrid analysis approach and other hybrid life cycle inventory techniques. This would capitalise on the data processing module created for the path exchange method. The SPA module would not be used in this case and the hybridisation module would need to include new classes and functions to accommodate other hybrid analysis techniques. This would enable a robust comparison of different hybridisation techniques using the same datasets, and would overcome the limitations of each.

In the longer term, progress in artificial intelligence may be harnessed to support the identification of corresponding input-output and process *nodes*. However, as shown by Miotto et al. (2017), the deep learning method used in many artificial intelligence models is often a black box, which means that the model would provide the prediction of an outcome with no explanation. In the context of research on hybrid LCA, transparency is key in justifying any given outcome. The reasoning behind decisions made in identifying corresponding input-output and process *nodes* must be transparent. In addition, there are currently no clear guidelines on the amount of information required to successfully train a deep neural network (Miotto et al. 2017), and these models perform best where large databases of information are available for training purposes. In the case of hybrid LCIs, it would most likely be necessary to manually conduct thousands of hybridisations to provide sufficient information for training a neural network. Using artificial intelligence to compile hybrid LCIs would therefore require the manual compilation of such inventories first in order to be able to effectively train neural networks. Subsequently, trained neural networks should ideally be able to provide, alongside the Path Exchange, the reasoning behind the choice of matching pathways. While future research will reveal if artificial intelligence might be able to automatically produce hybrid LCIs, the proposed model is set up to be easily combined with neural networks, if and when this becomes possible.

6. Conclusion

This paper has described the development of an automated model for the application of the Path Exchange hybrid LCI analysis method. The main purpose of this automation is to address current barriers to the wider use of hybrid LCI methods. By reducing the time and complexity associated with compiling a hybrid LCI it is hoped that hybrid analysis can be made more accessible to the LCA community. This study has demonstrated the compilation of a hybrid GHG emissions value for a basic construction product relating to Australian production. However, the strength of the model is its ability to be used to calculate hybrid values for any environmental flow for any variety of products in any region of the world, as long as the necessary data is available.

The intention is for the model to be made open-source so that LCA specialists can easily adapt the model to suit particular contexts and datasets. This is facilitated by the object-oriented programming that has been used, which enables greater flexibility and versatility. The open-source nature of the model will enable its collective development for the benefit of the wider LCA community. While the focus of the model is on automating the compilation of a PXC hybrid LCI, it could just as easily be used to automate the establishment of a pure process or input-output LCI, further streamlining various aspects of these more commonly used LCI approaches.

While the model provides a sophisticated and comprehensive attempt to automate hybrid LCI analysis, the model is not yet fully automated. However, rapidly emerging technologies such as artificial intelligence, will provide an opportunity to further automate various aspects of the model in the near future. Despite this, the current model is a critical first step towards improving widespread access to hybrid LCI methods, streamlining data compilation and enabling more comprehensive analyses. This is particularly useful for sectors such as the construction industry where substantial environmental improvements are needed to address critical global environmental issues.

Authors' contributions

RC and AS conceived the research and secured funding; RC, AS and PAB defined the model's functionalities; AS and PAB designed, developed, programmed and tested the model; PAB conducted the case study analysis; AS, RC and PAB wrote the paper.

Acknowledgements

This research was supported by the Australian Research Council's Discovery Projects funding scheme (project number DP150100962) and the Australian Research Council's Linkage Infrastructure, Equipment and Facilities funding scheme (project number LE160100066).

References

Alic D, Omanovic S, Giedrimas V (2016) Comparative analysis of functional and object-oriented programming. Paper presented at the 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, May 30th - June 3rd

Atlassian. (2017). SourceTree. Retrieved from https://www.atlassian.com/software/sourcetree

Augspurger T et al. (2017). Python Data Analysis Library. Retrieved from http://pandas.pydata.org/index.html

Baboulet O (2009) Path Exchange Method for Hybrid Life-Cycle Assessment Norwegian University of Science and Technology

Blengini GA, Di Carlo T (2010) The changing role of life cycle phases, subsystems and materials in the LCA of low energy buildings Energ Buildings 42:869-880 doi:10.1016/j.enbuild.2009.12.009

Bontinck P-A (2018) Summary of path exchange hybrid life cycle greenhouse gas emissions inventory for Plasterboard in Australia. Figshare. https://doi.org/10.6084/m9.figshare.5687770. Accessed February 23rd 2018

Bontinck P-A, Crawford RH, Stephan A (2017) Improving the Uptake of Hybrid Life Cycle Assessment in the Construction Industry Procedia Engineering 196:822-829 doi:https://doi.org/10.1016/j.proeng.2017.08.013

Crama Y, Defourny J, Gazon J (1984) Structural Decompostion of Multipliers in Input-Output or Social Accounting Matrix Analysis Economie Appliquee 37:215-222 doi:http://www.ismea.org/ISMEA/eapp.html

Crawford RH (2011) Life cycle assessment in the built environment. Spon Press, London

Crawford RH, Bartak EL, Stephan A, Jensen CA (2016) Evaluating the life cycle energy benefits of energy efficiency regulations for buildings Renew Sust Energ Rev 63:435-451 doi:http://dx.doi.org/10.1016/j.rser.2016.05.061

Crawford RH, Bontinck P-A, Stephan A, Wiedmann T (2017) Towards an automated approach for compiling hybrid life cycle inventories Procedia Eng 180:157-166 doi:doi: 10.1016/j.proeng.2017.04.175

Crawford RH, Bontinck P-A, Stephan A, Wiedmann T, Yu M (2018a) Hybrid life cycle inventory methods – a review J Clean Prod 172:1273-1288 doi:https://doi.org/10.1016/j.jclepro.2017.10.176

Crawford RH, Stephan A, Bontinck P-A (2018b) Improving the life cycle environmental performance of Australian construction projects. https://figshare.com/projects/Improving_the_life_cycle_environmental_performance_of_Australian_construction_projects/27433. Accessed 11th of May 2018

Defourny J, Thorbecke E (1984) Structural path analysis and multiplier decomposition within a social accounting matrix framework Economic Journal 94:111-136

Dixit MK (2017) Life cycle embodied energy analysis of residential buildings: A review of literature to investigate embodied energy parameters Renew Sust Energ Rev 79:390-413 doi:https://doi.org/10.1016/j.rser.2017.05.051

Dubois PF (2005) Maintaining correctness in scientific programs Computing in Science & Engineering 7:80-85 doi:10.1109/MCSE.2005.54

Eurostat (2008) Eurostat manual of supply, use and input-output tables. Office for Official Publications of the European Communities, Luxembourg

Gibon T, Schaubroeck T (2017) Lifting the fog on characteristics and limitations of hybrid LCA—a reply to "Does hybrid LCA with a complete system boundary yield adequate results for product promotion?" by Yi Yang (Int J Life Cycle Assess 22(3):456–406, doi:10.1007/s11367-016-1256-9 Int J Life Cycle Ass 22:1005-1008 doi:10.1007/s11367-017-1291-1

Grant T (2016) AusLCI Database Manual v1.26, 1.26 edn. Australian Life Cycle Assessment Society (ALCAS),

Hammond G, Jones C (2008) Inventory of carbon & energy: ICE. Sustainable Energy Research Team, Department of Mechanical Engineering, University of Bath, Bath, UK

Junnila S (2004) The environmental impact of an office building throughout its life cycle. PhD thesis, Helsinki University of Technology Construction Economics and Management

Lenzen M (2000) Errors in Conventional and Input-Output-based Life-Cycle Inventories J Ind Ecol 4:127-148 doi:10.1162/10881980052541981

Lenzen M, Crawford RH (2009) The Path Exchange Method for Hybrid LCA Environ Sci Technol 43:8251-8256

Lenzen M et al. (2014) Compiling and using input–output frameworks through collaborative virtual laboratories Science of the Total Environment 485-486:241-251 doi:10.1016/j.scitotenv.2014.03.062

Leontief W (1970) Environmental repercussions and the economic structure: an input-output approach Rev Econ Statistics 52:262-271

Lutz M (2013) Learning Python. Fifth edn. O'Reilly Media, Sebastopol, CA

Majeau-Bettez G. (2014). ecospold2matrix: A Python class for recasting Ecospold2 LCA datasets into Leontief matrix representations or supply and use tables. GitHub. Retrieved from https://github.com/majeau-bettez/allocation_construct

Majeau-Bettez G, Strømman AH, Hertwich EG (2011) Evaluation of process- and input-output-based life cycle inventory data with regard to truncation and aggregation issues Environ Sci Technol 45:10170-10177

Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2017) Deep learning for healthcare: review, opportunities and challenges Briefings In Bioinformatics doi:10.1093/bib/bbx044

Oregi X, Hernandez P, Gazulla C, Isasa M (2015) Integrating Simplified and Full Life Cycle Approaches in Decision Making for Building Energy Refurbishment: Benefits and Barriers Buildings 5:354-380

Pauliuk S, Majeau-Bettez G, Mutel CL, Steubing B, Stadler K (2015) Lifting Industrial Ecology Modeling to a New Level of Quality and Transparency: A Call for More Transparent Publications and a Collaborative Open Source Software Framework J Ind Ecol 19:937-949 doi:10.1111/jiec.12316

Pomponi F, Lenzen M (2018) Hybrid life cycle assessment (LCA) will likely yield more accurate results than process-based LCA J Clean Prod 176:210-215 doi:https://doi.org/10.1016/j.jclepro.2017.12.119

PRé Sustainability. (2018). Simapro (Version 8.4). Retrieved from https://simapro.com/

Python Software Foundation. (2017). Python programming language - official website (Version 3.4). Retrieved from www.python.org

Rauf A, Crawford RH (2015) Building Service Life and its Effect on the Life Cycle Embodied Energy of Buildings Energy 79:140-148 doi:http://dx.doi.org/10.1016/j.energy.2014.10.093

Säynäjoki A, Heinonen J, Junnila S, Horvath A (2017) Can life-cycle assessment produce reliable policy guidelines in the building sector? Environ Res Lett 12:013001 doi:https://doi.org/10.1088/1748-9326/aa54ee

Schaubroeck T, Gibon T (2017) Outlining reasons to apply hybrid LCA—a reply to "rethinking system boundary in LCA" by Yi Yang (2017) Int J Life Cycle Ass 22:1012-1013 doi:10.1007/s11367-017-1311-1

Stephan A (2018) Walkthrough the Path Exchange Hybrid Analysis Graphical User Interface. Figshare. https://doi.org/10.6084/m9.figshare.5687773. Accessed February 23rd 2018

Stephan A, Stephan L (2014) Reducing the total life cycle energy demand of recent residential buildings in Lebanon Energy 74:618-637 doi:http://dx.doi.org/10.1016/j.energy.2014.07.028

Suh S, Huppes G (2000) Gearing input-output model to LCA, part I: General framework for hybrid approach. CML, Leiden University, Leiden, The Netherlands

Suh S, Huppes G (2005) Methods for Life Cycle Inventory of a Product J Clean Prod 13:687-697 doi:10.1016/j.jclepro.2003.04.001

Suh S et al. (2004) System Boundary Selection in Life-Cycle Inventories Using Hybrid Approaches Environ Sci Technol 38:657-664

Suh S, Weidema B, Schmidt JH, Heijungs R (2010) Generalized Make and Use Framework for Allocation in Life Cycle Assessment J Ind Ecol 14:335-353 doi:http://dx.doi.org/10.1111/j.1530-9290.2010.00235.x

Tettey UYA, Dodoo A, Gustavsson L (2014) Effects of different insulation materials on primary energy and CO2 emission of a multi-storey residential building Energ Buildings 82:369-377 doi:http://dx.doi.org/10.1016/j.enbuild.2014.07.009

Treloar GJ (1997) Extracting embodied energy paths from input-output tables: towards an input-output-based hybrid energy analysis method Econ Systems Res 9:375-391

Treloar GJ (1998) A comprehensive embodied energy analysis framework. Ph.D. Thesis, Deakin University

van Rossum G, Warsaw B, Coghlan N (2018) PEP 8 -- Style Guide for Python Code. https://www.python.org/dev/peps/pep-0008/. Accessed 11th May 2018

WU Global Material Flows Database (2017) Vienna University of Economics and Business. www.materialflows.net.

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 Ass 21:1218-1230 doi:https://doi.org/10.1007/s11367-016-1087-8

Wilson G et al. (2014) Best Practices for Scientific Computing PLoS Biology 12:1-7 doi:10.1371/journal.pbio.1001745

World Bank Group (2017) Population, total.

wxPython (2018) Welcome to wxPython. http://www.wxpython.org/. Accessed February 21st 2018

Yang Y (2017a) Does hybrid LCA with a complete system boundary yield adequate results for product promotion? Int J Life Cycle Ass 22:456-460 doi:10.1007/s11367-016-1256-9

Yang Y (2017b) Rethinking system boundary in LCA—reply to "Lifting the fog on the characteristics and limitations of hybrid LCA" by Thomas Gibon and Thomas Schaubroeck (2017) Int J Life Cycle Ass 22:1009-1011 doi:10.1007/s11367-017-1295-x

Yang Y, Heijungs R, Brandão M (2017) Hybrid life cycle assessment (LCA) does not necessarily yield more accurate results than process-based LCA J Clean Prod 150:237-242 doi:https://doi.org/10.1016/j.jclepro.2017.03.006

Figures

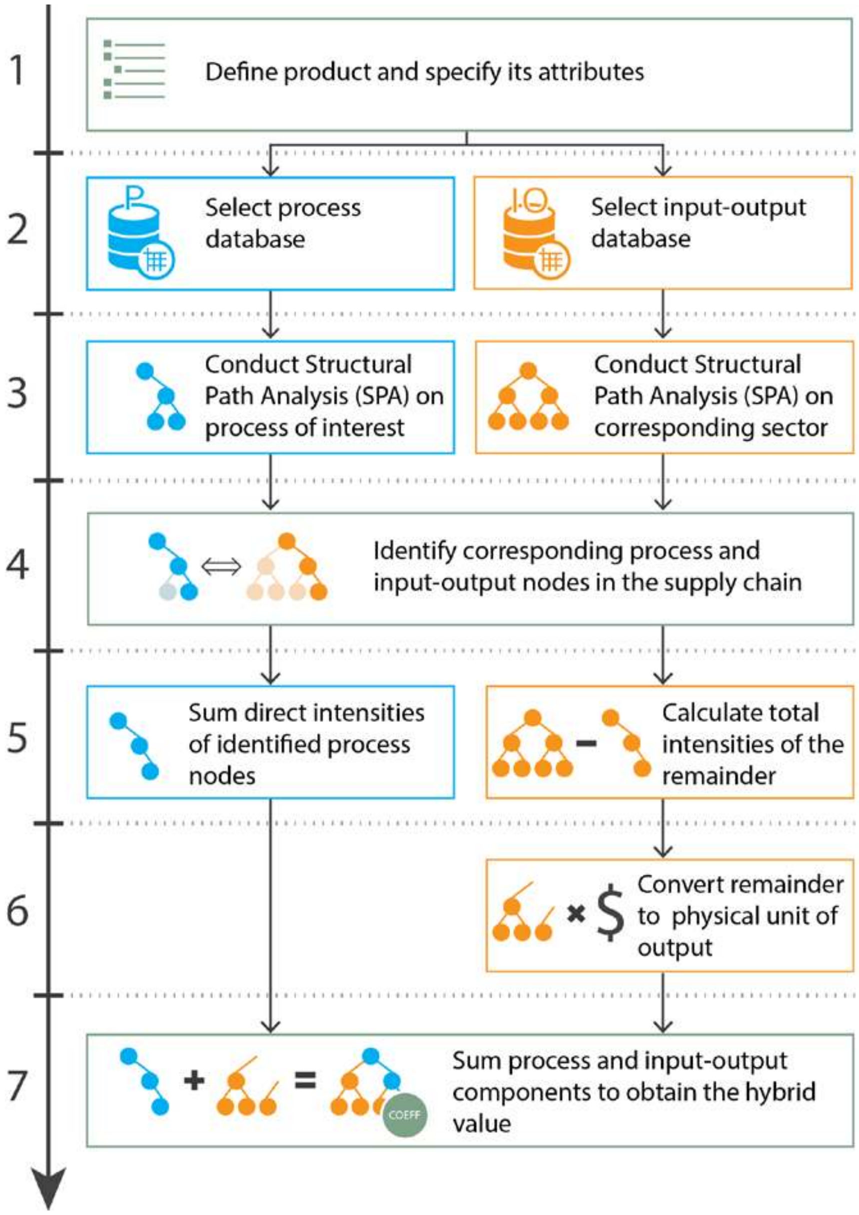

1
Define product and specify its attributes
2
Select process database
Select input-output database
3
Conduct Structural Path Analysis (SPA) on process of interest
Conduct Structural Path Analysis (SPA) on corresponding sector
4
Identify corresponding process and input-output nodes in the supply chain
5
Sum direct intensities of identified process nodes
Calculate total intensities of the remainder
6
Convert remainder to physical unit of output
7
Sum process and input-output components to obtain the hybrid value
COEFF

Figure 1 Step-by-step description of the Path Exchange hybrid analysis method.

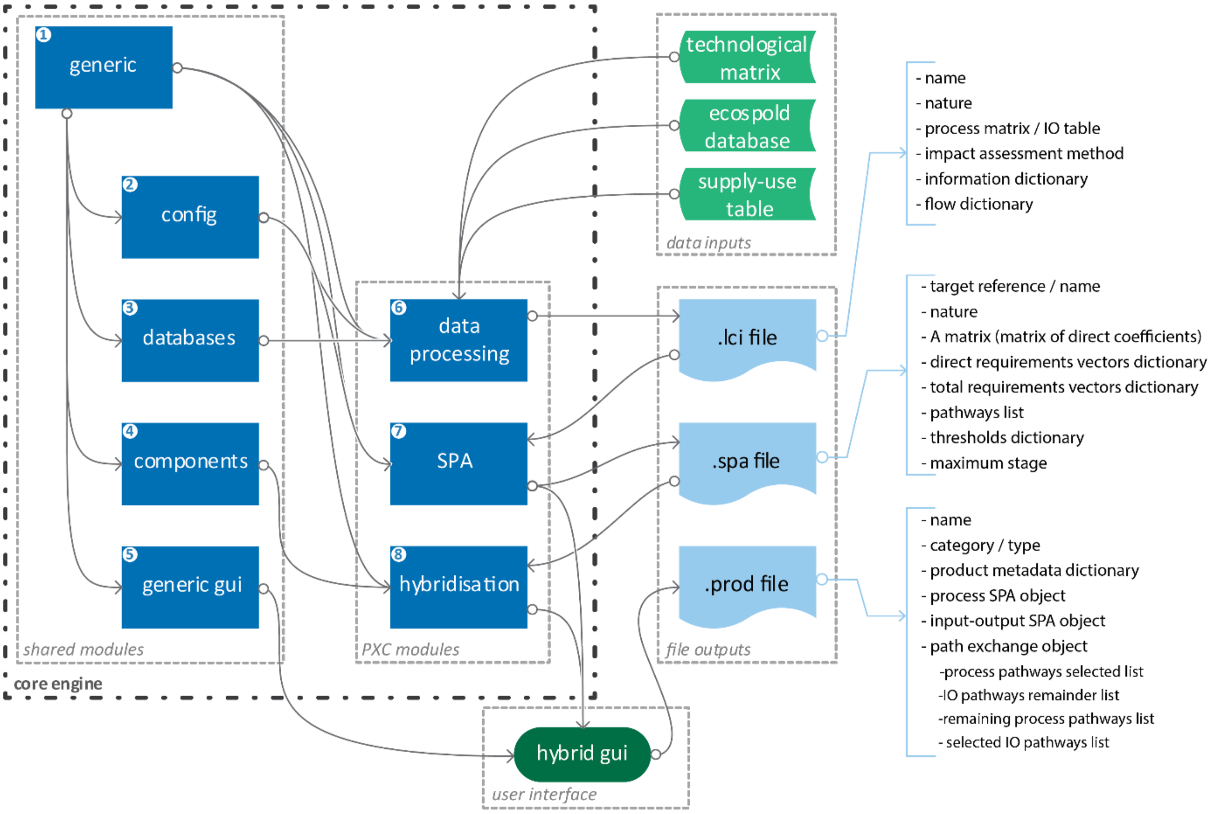


Figure 2 General wire diagram of the automated PXC hybrid model.

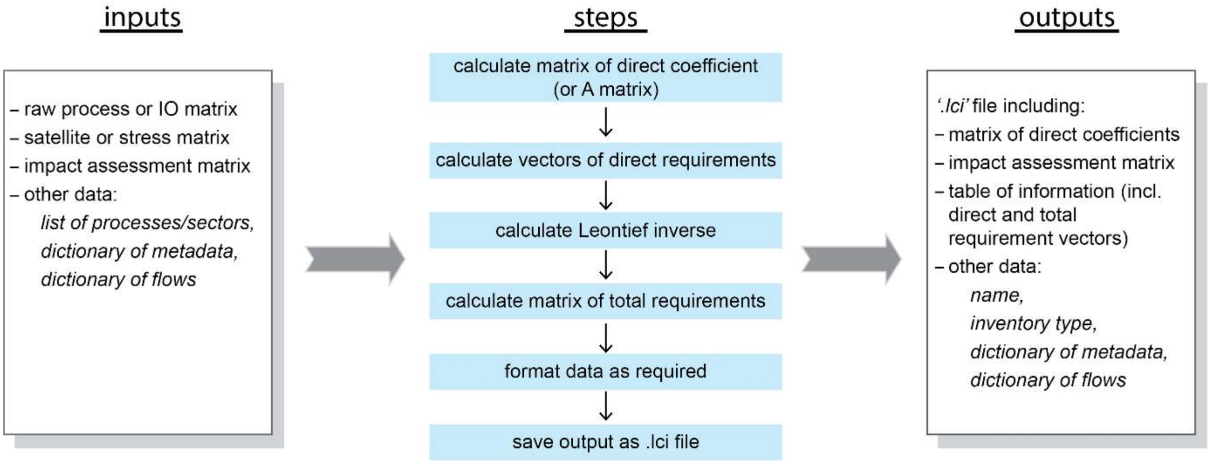
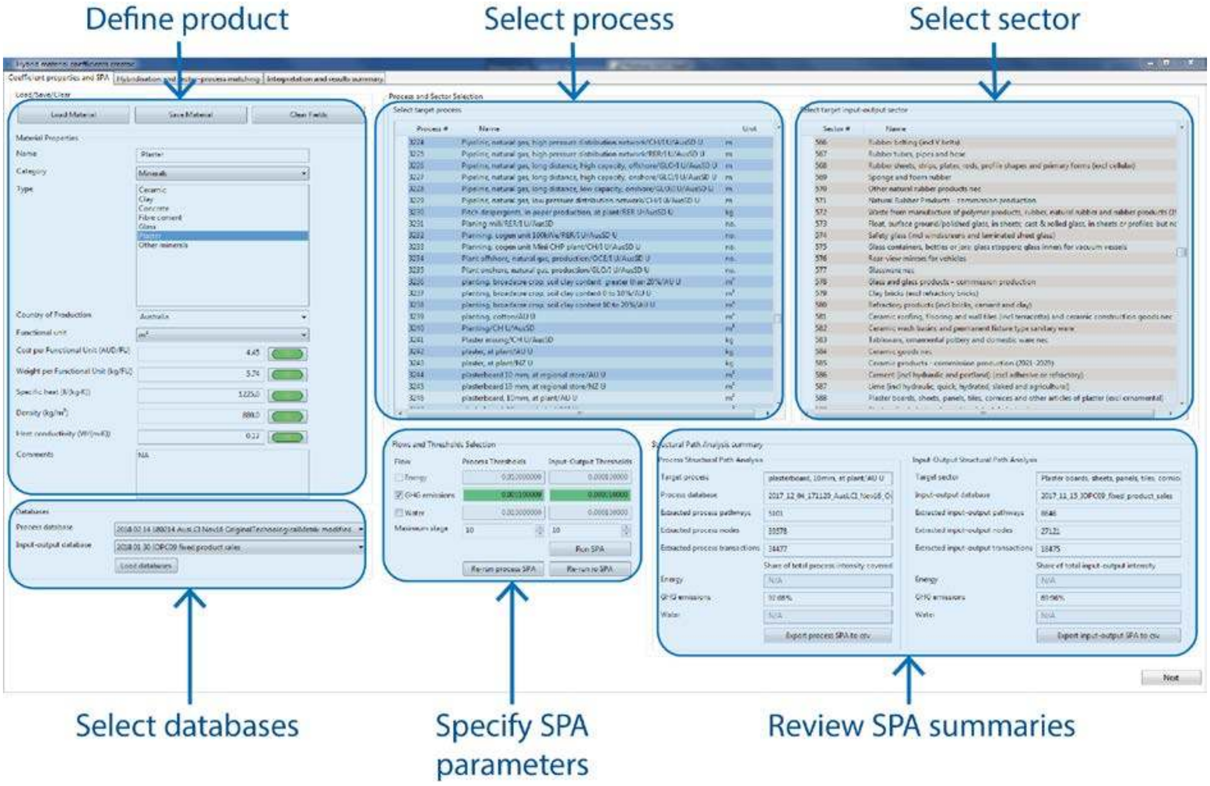

inputs
– raw process or IO matrix
– satellite or stress matrix
– impact assessment matrix
– other data:
list of processes/sectors,
dictionary of metadata,
dictionary of flows
steps
calculate matrix of direct coefficient (or A matrix)
calculate vectors of direct requirements
calculate Leontief inverse
calculate matrix of total requirements
format data as required
save output as .lci file
outputs
'.lci' file including:
– matrix of direct coefficients
– impact assessment matrix
– table of information (incl. direct and total requirement vectors)
– other data:
name,
inventory type,
dictionary of metadata,
dictionary of flows

Figure 3 Schematic overview of the data processing module.



Hybrid material coefficients creator

Coefficient properties and SPA | Hybridisation and sector-process matching | Integrated and results summary

Load/Save/Clear

Load Material | Save Material | Clear Fields

Material Properties

Name	Plaster
Category	Minerals
Type	Ceramic Clay Concrete Fibre cement Glass Plaster Other minerals
Country of Production	Australia
Functional unit	m²
Cost per Functional Unit (AUD/FU)	4.45
Weight per Functional Unit (kg/FU)	5.74
Specific heat (J/(kg·K))	1225.0
Density (kg/m³)	880.0
Heat conductivity (W/(m·K))	0.22
Comments	NA

Databases

Process database	2018 02 14 180014 AusLCI Nov16 OriginalTechnologicalMatrix modified
Input-output database	2018 01 30 IOPC09 fixed product sales

Load databases

Process and Sector Selection

Select target process

Process #	Name	Unit
3224	Pipeline, natural gas, high pressure distribution network/CH/I U/AusSD U	m
3225	Pipeline, natural gas, high pressure distribution network/RER/I U/AusSD U	m
3226	Pipeline, natural gas, long distance, high capacity, offshore/GLO/I U/AusSD U	m
3227	Pipeline, natural gas, long distance, high capacity, onshore/GLO/I U/AusSD U	m
3228	Pipeline, natural gas, long distance, low capacity, onshore/GLO/I U/AusSD U	m
3229	Pipeline, natural gas, low pressure distribution network/CH/I U/AusSD U	m
3230	Pitch despergents, in paper production, at plant/RER U/AusSD U	kg
3231	Planing mill/RER/I U/AusSD	no.
3232	Planning, cogen unit 160kWe/RER/I U/AusSD U	no.
3233	Planning, cogen unit Mini CHP plant/CH/I U/AusSD U	no.
3234	Plant offshore, natural gas, production/OCE/I U/AusSD U	no.
3235	Plant onshore, natural gas, production/GLO/I U/AusSD U	no.
3236	planting, broadacre crop, soil clay content greater than 20%/AU U	m²
3237	planting, broadacre crop, soil clay content 0 to 10%/AU U	m²
3238	planting, broadacre crop, soil clay content 10 to 20%/AU U	m²
3239	planting, cotton/AU U	m²
3240	Planting/CH U/AusSD	m²
3241	Plaster mixing/CH U/AusSD	kg
3242	plaster, at plant/AU U	kg
3243	plaster, at plant/NZ U	kg
3244	plasterboard 10 mm, at regional store/AU U	m²
3245	plasterboard 10 mm, at regional store/NZ U	m²
3246	plasterboard, 10mm, at plant/AU U	m²

Select target input-output sector

Sector #	Name
566	Rubber belting (incl V belts)
567	Rubber tubes, pipes and hose
568	Rubber sheets, strips, plates, rods, profile shapes and primary forms (excl cellular)
569	Sponge and foam rubber
570	Other natural rubber products nec
571	Natural Rubber Products - commission production
572	Waste from manufacture of polymer products, rubber, natural rubber and rubber products (i
573	Float, surface ground/polished glass, in sheets; cast & rolled glass, in sheets or profiles but nc
574	Safety glass (incl windscreens and laminated sheet glass)
575	Glass containers, bottles or jars; glass stoppers; glass inners for vacuum vessels
576	Rear view mirrors for vehicles
577	Glassware nec
578	Glass and glass products - commission production
579	Clay bricks (excl refractory bricks)
580	Refractory products (excl bricks, cement and clay)
581	Ceramic roofing, flooring and wall tiles (incl terracotta) and ceramic construction goods nec
582	Ceramic wash basins and permanent fixture type sanitary ware
583	Tableware, ornamental pottery and domestic ware nec
584	Ceramic goods nec
585	Ceramic products - commission production (2021-2029)
586	Cement (incl hydraulic and portland); (excl adhesive or refractory)
587	Lime (incl hydraulic, quick, hydrated, slaked and agricultural)
588	Plaster boards, sheets, panels, tiles, cornices and other articles of plaster (excl ornamental)

Flows and Thresholds Selection

Flow	Process Thresholds	Input-Output Thresholds
□ Energy	0.001000000	0.000010000
☑ GHG emissions	0.001000000	0.000010000
□ Water	0.001000000	0.000010000
Maximum stage	10	10

Run SPA

Re-run process SPA | Re-run io SPA

Structural Path Analysis summary

Process Structural Path Analysis

Target process	plasterboard, 10mm, at plant/AU U
Process database	2017_12_04_171120_AusLCI_Nov16_O
Extracted process pathways	5101
Extracted process nodes	33578
Extracted process transactions	34477
	Share of total process intensity covered
Energy	N/A
GHG emissions	92.68%
Water	N/A

Export process SPA to csv

Input-Output Structural Path Analysis

Target sector	Plaster boards, sheets, panels, tiles, cornic
Input-output database	2017_11_15_IOPC09_fixed_product_sales
Extracted input-output pathways	8646
Extracted input-output nodes	27121
Extracted input-output transactions	18475
	Share of total input-output intensity
Energy	N/A
GHG emissions	89.96%
Water	N/A

Export input-output SPA to csv

Next

Figure 4 Overview of the first tab of the model’s graphical user interface.

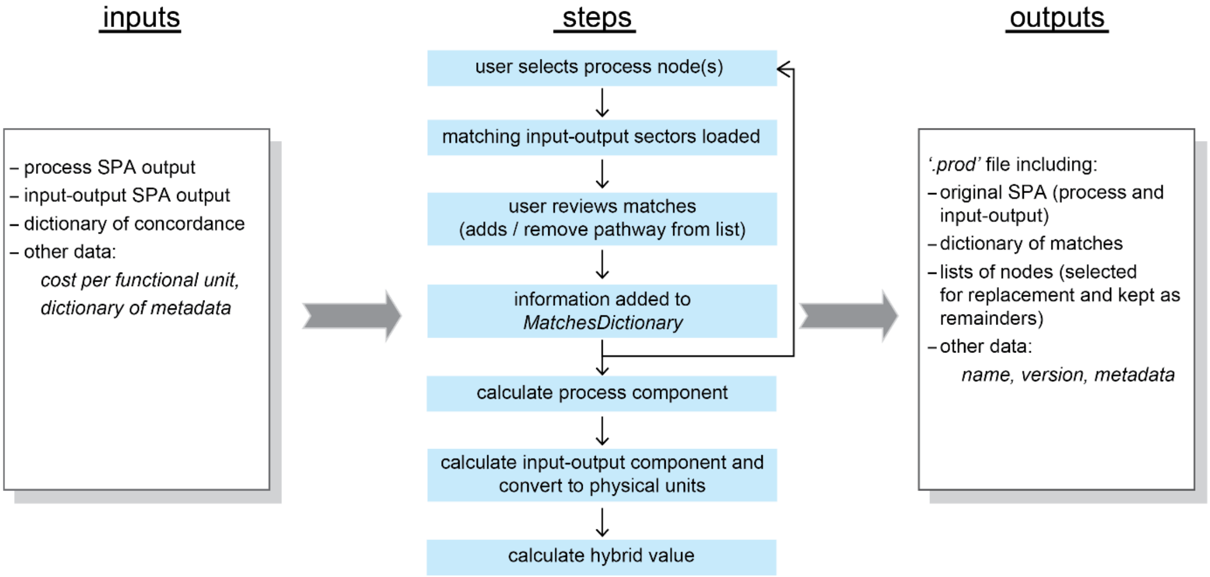
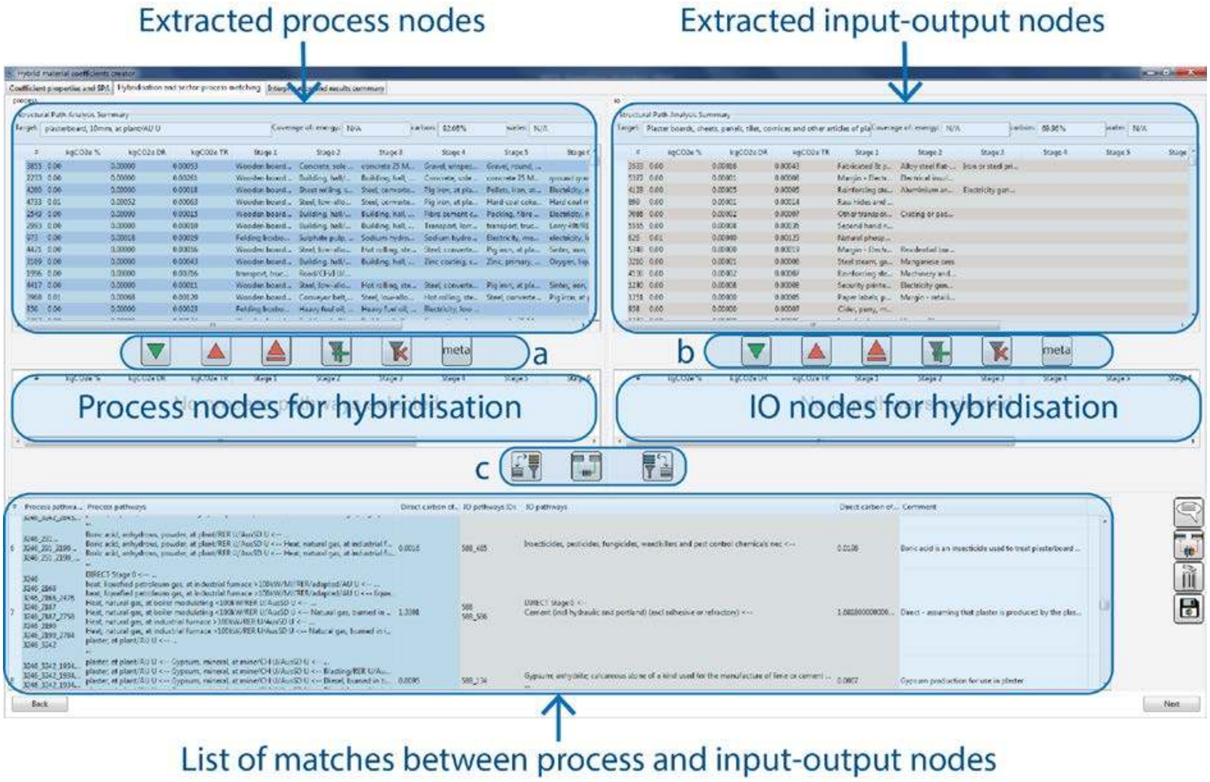

inputs
steps
outputs
– process SPA output
– input-output SPA output
– dictionary of concordance
– other data:
cost per functional unit, dictionary of metadata
user selects process node(s)
matching input-output sectors loaded
user reviews matches (adds / remove pathway from list)
information added to MatchesDictionary
calculate process component
calculate input-output component and convert to physical units
calculate hybrid value
'.prod' file including:
– original SPA (process and input-output)
– dictionary of matches
– lists of nodes (selected for replacement and kept as remainders)
– other data:
name, version, metadata

Figure 5 Schematic overview of the hybridisation module.


Extracted process nodes
Extracted input-output nodes
Hybrid material coefficients creator
Coefficient properties and SPA
Hybridisation and sector process matching
Interpretation and results summary
process
io
a
b
meta
Process nodes for hybridisation
IO nodes for hybridisation
c
Back
Next
List of matches between process and input-output nodes

Structural Path Analysis Summary

Target: plasterboard, 10mm, at plant/AU U Coverage of: energy: N/A carbon: 82.68% water: N/A

#	kgCO2e %	kgCO2e DR	kgCO2e TR	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
3855	0.06	0.00000	0.00053	Wooden board...	Concrete, sole ...	concrete 25 M...	Gravel, unspec...	Gravel, round, ...	
3713	0.06	0.00000	0.00261	Wooden board...	Building, hall/...	Building, hall, ...	Concrete, sole ...	concrete 25 M...	ground gra
4280	0.06	0.00000	0.00018	Wooden board...	Sheet rolling, s...	Steel, converte...	Pig iron, at pla...	Pellets, iron, at...	Electricity, m
4733	0.01	0.00052	0.00063	Wooden board...	Steel, low-allo...	Steel, converte...	Pig iron, at pla...	Hard coal coke...	Hard coal m
2549	0.06	0.00000	0.00015	Wooden board...	Building, hall/...	Building, hall, ...	Fibre cement c...	Packing, fibre ...	Electricity, m
2993	0.06	0.00000	0.00010	Wooden board...	Building, hall/...	Building, hall, ...	Transport, lorr...	transport, truc...	Lorry 40t/RE
973	0.06	0.00018	0.00029	Folding boxbo...	Sulphate pulp, ...	Sodium hydro...	Sodium hydro...	Electricity, me...	electricity, h
4471	0.06	0.00000	0.00016	Wooden board...	Steel, low-allo...	Hot rolling, ste...	Steel, converte...	Pig iron, at pla...	Sinter, iron,
3169	0.06	0.00000	0.00043	Wooden board...	Building, hall/...	Building, hall, ...	Zinc coating, c...	Zinc, primary, ...	Oxygen, liq
1996	0.06	0.00000	0.00756	transport, truc...	Road/CH/I U/...				
4417	0.06	0.00000	0.00011	Wooden board...	Steel, low-allo...	Hot rolling, ste...	Steel, converte...	Pig iron, at pla...	Sinter, iron,
3968	0.01	0.00068	0.00120	Wooden board...	Conveyor belt,...	Steel, low-allo...	Hot rolling, ste...	Steel, converte...	Pig iron, at
850	0.06	0.00000	0.00023	Folding boxbo...	Heavy fuel oil, ...	Heavy fuel oil, ...	Electricity, low ...		

Structural Path Analysis Summary

Target: Plaster boards, sheets, panels, tiles, cornices and other articles of pla Coverage of: energy: N/A carbon: 69.86% water: N/A

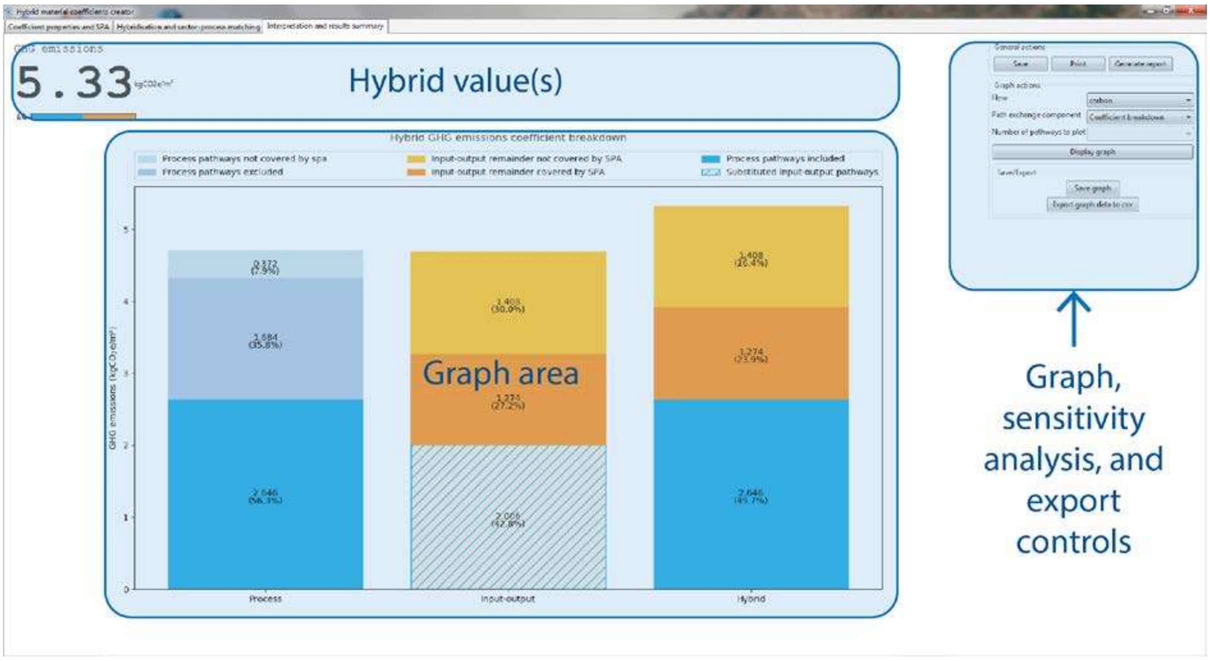
#	kgCO2e %	kgCO2e DR	kgCO2e TR	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage
2553	0.60	0.00006	0.00043	Fabricated St p...	Alloy steel flat-...	Iron or steel pri...			
5373	0.60	0.00001	0.00068	Margin - Electr...	Electrical insul...				
4128	0.60	0.00005	0.00085	Reinforcing ste...	Aluminium an...	Electricity gen...			
890	0.60	0.00001	0.00014	Raw hides and ...					
3698	0.60	0.00002	0.00087	Other transpor...	Crating or pac...				
5555	0.60	0.00008	0.00035	Second hand r...					
625	0.61	0.00089	0.00125	Natural phosp...					
5148	0.60	0.00008	0.00019	Margin - Distrib...	Residential ...				
3210	0.60	0.00001	0.00086	Steel steam, ga...	Manganese ores				
4130	0.60	0.00002	0.00087	Reinforcing ste...	Machinery and ...				
3240	0.60	0.00008	0.00088	Security printe...	Electricity gen...				
3158	0.60	0.00000	0.00085	Paper labels, p...	Margin - retail...				
858	0.60	0.00000	0.00087	Cider, perry, m...					

#	kgCO2e %	kgCO2e DR	kgCO2e TR	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6

#	kgCO2e %	kgCO2e DR	kgCO2e TR	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6

#	Process pathwa...	Process pathways	Direct carbon of...	IO pathways IDs	IO pathways	Direct carbon of...	Comment
6	3246_251... 3246_251_2199... 3246_251_2198...	Boric acid, anhydrous, powder, at plant/RER U/AusSD U <-- ... Boric acid, anhydrous, powder, at plant/RER U/AusSD U <-- Heat, natural gas, at industrial f... Boric acid, anhydrous, powder, at plant/RER U/AusSD U <-- Heat, natural gas, at industrial f...	0.0016	588_465	Insecticides, pesticides, fungicides, weedkillers and pest control chemicals nec <--	0.0108	Boric acid is an insecticide used to treat plasterboard ...
7	3246 3246_2869 3246_2869_2476 3246_2867 3246_2867_2758 3246_2890 3246_2890_2784 3246_3242	DIRECT Stage 0 <-- ... heat, liquefied petroleum gas, at industrial furnace >100kW/MJ/RER/adapted/AU U <-- ... heat, liquefied petroleum gas, at industrial furnace >100kW/MJ/RER/adapted/AU U <-- Expa... Heat, natural gas, at boiler modulating <100kW/RER U/AusSD U <-- ... Heat, natural gas, at boiler modulating <100kW/RER U/AusSD U <-- Natural gas, burned in ... Heat, natural gas, at industrial furnace >100kW/RER U/AusSD U <-- ... Heat, natural gas, at industrial furnace >100kW/RER U/AusSD U <-- Natural gas, burned in i... plaster, at plant/AU U <-- ...	1.3398	588 588_586	DIRECT Stage 0 <-- Cement (incl hydraulic and portland) (excl adhesive or refractory) <--	1.000000000000...	Direct - assuming that plaster is produced by the plas...
8	3246_3242_1934... 3246_3242_1934... 3246_3242_1934...	plaster, at plant/AU U <-- Gypsum, mineral, at mine/CH U/AusSD U <-- ... plaster, at plant/AU U <-- Gypsum, mineral, at mine/CH U/AusSD U <-- Blasting/RER U/Au... plaster, at plant/AU U <-- Gypsum, mineral, at mine/CH U/AusSD U <-- Diesel, burned in ...	0.0095	588_134	Gypsum; anhydrite; calcareous stone of a kind used for the manufacture of lime or cement ...	0.0872	Gypsum production for use in plaster

Figure 6: Annotated screenshot of the hybridisation tab of the graphical user interface. Note: a: controls to select, filter and access metadata of extracted process *nodes*; b: controls to select, filter and access metadata of extracted input-output *nodes*; and c: controls to automatically find input-output or process *nodes* that match the current selection.


Hybrid material coefficients creator
Coefficient properties and SPA
Hybridisation and sector-process matching
Interpretation and results summary
GHG emissions
5.33 kgCO2e/m²
Hybrid value(s)
Hybrid GHG emissions coefficient breakdown
Process pathways not covered by spa
Process pathways excluded
Input-output remainder not covered by SPA
Input-output remainder covered by SPA
Process pathways included
Substituted input-output pathways
GHG emissions (kgCO₂-e/m²)
0
1
2
3
4
5
0.372 (7.9%)
1.684 (35.8%)
2.646 (56.3%)
1.408 (30.0%)
1.274 (27.2%)
2.006 (42.8%)
1.408 (26.4%)
1.274 (23.9%)
2.646 (49.7%)
Process
Input-output
Hybrid
Graph area
General actions
Save
Print
Generate report
Graph actions
Display graph
Save/Export
Save graph
Export graph data to csv
Graph, sensitivity analysis, and export controls

Flow: carbon

Path exchange component: Coefficient breakdown

Number of pathways to plot:

Figure 7 Annotated screenshot of the interpretation tab of the graphical user interface.

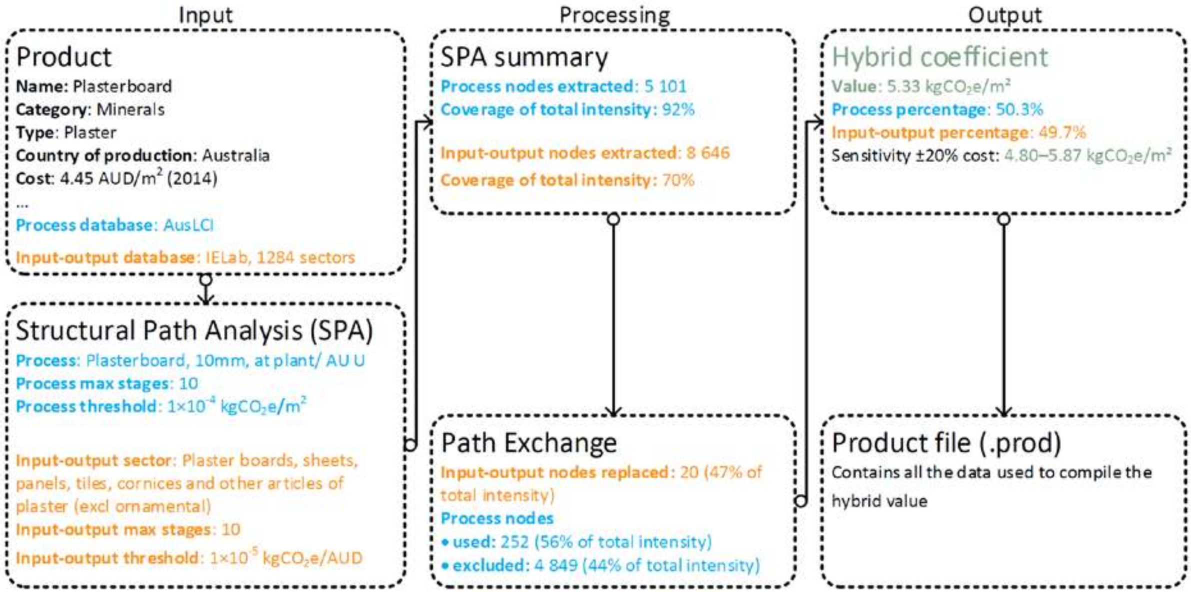

Input
Product
Name: Plasterboard
Category: Minerals
Type: Plaster
Country of production: Australia
Cost: 4.45 AUD/m² (2014)
...
Process database: AusLCI
Input-output database: IELab, 1284 sectors
Structural Path Analysis (SPA)
Process: Plasterboard, 10mm, at plant/ AU U
Process max stages: 10
Process threshold: 1×10⁻⁴ kgCO₂e/m²
Input-output sector: Plaster boards, sheets, panels, tiles, cornices and other articles of plaster (excl ornamental)
Input-output max stages: 10
Input-output threshold: 1×10⁻⁵ kgCO₂e/AUD
Processing
SPA summary
Process nodes extracted: 5 101
Coverage of total intensity: 92%
Input-output nodes extracted: 8 646
Coverage of total intensity: 70%
Path Exchange
Input-output nodes replaced: 20 (47% of total intensity)
Process nodes
• used: 252 (56% of total intensity)
• excluded: 4 849 (44% of total intensity)
Output
Hybrid coefficient
Value: 5.33 kgCO₂e/m²
Process percentage: 50.3%
Input-output percentage: 49.7%
Sensitivity ±20% cost: 4.80–5.87 kgCO₂e/m²
Product file (.prod)
Contains all the data used to compile the hybrid value

Figure 8 Summary of the model inputs, processing and outputs to generate a hybrid embodied GHG emissions value for a square metre of 10 mm plasterboard in Australia.