Green Chemistry

ROYAL SOCIETY OF CHEMISTRY

CRITICAL REVIEW

View Article Online
View Journal | View Issue

Cite this: *Green Chem.*, 2024, **26**, 9554

Procedural life cycle inventory of chemical products at laboratory and pilot scale: a compendium†

Daniele Cespi [a,b]

Life Cycle Assessment (LCA) is widely acknowledged by academia and industry as a key tool for promoting environmental sustainability within the field of green chemistry. However, certain barriers hinder its straightforward applicability, primarily stemming from data unavailability when the target reaction falls outside the direct or indirect control of LCA practitioners. Several methodologies have been proposed over the years to address the data gap in terms of mass, energy, catalyst, emissions, recovery, *etc.* These have been compiled into a compendium aimed at providing comprehensive guidance for practitioners in overcoming this challenge. This procedural life cycle inventory aims to facilitate the adoption of LCA by ensuring that key environmental steps, such as energy consumption, are not overlooked, and that mass balances are complete. The methodology is presented through a case study focusing on bio-based maleic anhydride.

Received 20th March 2024,
Accepted 2nd August 2024

DOI: 10.1039/d4gc01372g

rsc.li/greenchem

Introduction

The application of Life Cycle Assessment (LCA)[1,2] to chemical products is right now a consolidated approach within academia and industry.[3] Supporting research and development, ecolabelling and communication of the environmental results among stakeholders represent some major applications.[4–14]

As is well known, the application of the life cycle approach in the chemical sector is currently crucial. Understanding the major environmental hotspots in chemical synthesis plays a pivotal role in defining the right strategy to promote a zero-emission chemical industry. The chemistry sector faces various environmental challenges, including the use of hazardous reagents and auxiliaries, the release of harmful substances, the production of massive amounts of waste, and significant energy consumption. Additionally, the sector's renewability is far from being fully achieved, since still dominated by the usage of fossil resources. Since 1990, the European chemical industry has reduced greenhouse gas emissions by over 145 million tonnes CO_2eq. although the annual emissions remain high, at around 125 Mt CO_2eq. in 2021.[25] Additionally, chemical waste has decreased by nearly one-third since 2007, amounting to around 8 Mt. Accidental pollutant releases have dropped by at least 40%, and the emission of water pollutants has nearly halved, with total organic carbon (TOC) emissions around 13kt LCA serves as a scientific and technical methodology to support innovation and plan future investments to mitigate environmental footprints and costs. The savings can then be reinvested in research and development, analysing plant inefficiencies (*e.g.*, energy audits), and improving performance through the installation of the best available techniques.

Several sectors are involved, directly and not, due to the unavoidable use of different chemical substances in our day life and the recognition of the LCA as assessment methodology for some products certificate.[15–17] In addition, green chemistry community[18] encourages the adoption of life cycle metrics that goes beyond resources efficiencies (*e.g.*, mass yield, *E*-factor, atom economy, *etc.*), to cover potential impacts (direct and indirect) on human health, resources consumption and ecosystem quality of syntheses.

Around ten years ago Kralisch *et al.*[19] published a tutorial review on the issue, with the aim of supporting researchers in the field of green chemistry to get familiar with the methodology.

However, the adoption of the LCA approach still remains nothing than easy, several barriers and limits exist. First of all, the databases population. Around 500 different substances are estimated to be included within open source and commercial databases,[20] numbers not suitable to cover the sector vastness.

[a]*Department of Industrial Chemistry "Toso Montanari", University of Bologna, Via Piero Gobetti 85, 40136 Bologna, BO, Italy. E-mail: daniele.cespi2@unibo.it*
[b]*Center for Chemical Catalysis – C3, Alma Mater Studiorum Università di Bologna, Via Piero Gobetti 85, 40136 Bologna, BO, Italy*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4gc01372g

Among 85 000–100 000 of molecules are available on market and registered, the majority bulk and intermediate chemicals.[20,21] Such limitation is mainly due to the corporates know how, which restricts sharing information on production processes. As a consequence, primary data to fill dataset are not always available. The same situation, even if more complex, could be encountered in the case of new processes and emerging technologies: very limited data availability and high degree of uncertainty.[22,23] Second, the lack of knowledge: LCA practitioners not necessary hold a basic chemical background. Competences gained during several master/PhD degree courses (such as chemistry, industrial chemistry or chemical engineering) are essential to understand the process and its components, the nature of the precursors and of the waste streams, as well as to identify the economic background necessary for completing the inventory (*e.g.*, allocation rules). Knowledge of chemistry and chemical transformation also add credibility to the study.[23] In addition, in this case LCA deputies are intended being able to read patents and extract the requested information necessary for the purpose. Together with those barriers, several limitations for achieving a complete inventory of chemical substances could be identified. Among these the first is represented by the energy requirements, most of the times secreted by companies since considered sensitive data. Therefore, the difficulties to include primary data induces researchers to omit energy flows leading to rough LCA models, although the importance of such theme. Recent analysis conducted by eia[24] revealed how the chemical industry within the United States is responsible for the 37% of the whole industrial consumption (7141 trillion btu on 19 436 trillion btu in 2021). In the European chemical industry energy requirements have the same pivotal role. In fact, despite the reduction of 21% in 30 years, the requirements averaged 51 million of tonnes of oil equivalent in 2019.[25] Similar difficulties are identified for the catalytic systems, whose amount per desired product, the energy requirements for assembly and regeneration are often not accessible. In addition, difficulties in completing the life cycle inventory are sometimes related to the lack of precursors (time analysis extension with higher costs), none data on the releases (emissions and waste) as well as on the end-of-life stages (the latter not always necessary, since cradle-to-gate boundaries could be enough for intermediate chemicals). In this article, a step-by-step procedural life cycle inventory for chemicals is proposed, by assuming the perspective of a standard LCA (level 3, comprehensive) as described in literature.[23] The aim is to share a common approach, which takes into account the main findings from previous literature by proposing an approach that can be used as a compendium to support practitioners in simulating the production at a laboratory or pilot scale when the reaction is not controlled (directly or indirectly). The production of maleic anhydride from biomass was selected as a reference case study to present the approach of life cycle inventory improvement, given the importance worldwide of the building block and the increasing interest in the bio-based sector.

Materials and methods

As described above, the major criticality in assessing life cycle studies at laboratory and pilot scale of chemical products and reactions are data availability. Among these, the minimal information requested to cover the cradle-to-gate pathway for a chemical compound like the reaction efficiency (conversion, selectivity and yield), stoichiometry, releases (emissions and waste) and energy consumption are sometimes unavailable. Fig. 1 describes two extreme situations that the practitioners have to face to complete LCIs. Case A when the reaction is under direct or indirect control of the institution interested in performing the LCA. On the other hand, in the Case B none available possibilities of running and measuring the reaction parameters occur. The alternative A implies the reaction is conducted internally or within an external laboratory with which a collaboration is in place. In this case most of all the foreground‡ data derives from direct experiments and calculation. Therefore, they are classified as primary. At least a two-members team is necessary, with LCA skills and competencies of synthetic and analytical chemistry. Otherwise, B represents the situation in which the LCA simulation is necessary, but the reaction is not under control and cannot be investigated with direct experiments. This scenario occurs when the target substance originates from a supplier, from a company's viewpoint, or when there is a curiosity to explore its environmental characteristics due to research and development motives, such as academic investigation. In these cases, none primary data are available; information can be only estimated by literature and recalculation. In fact, even if software simulation is considered one of the more favoured method to complete the LCIs[26,27] it is challenging to apply in this instance due to the lack of control over the reaction. A process simulation analysis based on software engineering (also called techno-economic analysis) requires a minimal set of data to be fulfilled (*e.g.*, type of reactor, kinetics equations, reaction conditions, *etc.*). Therefore, without the access to such information the compilation seems not feasible at all. In addition, the simulation requires expertise and a dedicated budget, not always available. Hence, a procedural approach is outlined here, aimed at providing guidance for LCA experts with a chemical background, facilitating the completion of the LCI.

Mass balance (including emissions)

As a general rule of thumb each LCI has to respects the Lavoisier principle[28] without any material lost. Therefore, the practitioner should know and define the chemical reaction involved in the synthesis of the target molecule (under study), the type and nature of the reagents and co-products, as well as the reaction conditions. First step is knowing the reaction stoichiometry. A possibility is to extrapolate such information

‡ Foreground system consists of processes which are under the control of the decision-maker, in accordance with R. Frischknecht, Life cycle inventory analysis for decision-making. Scope-dependent inventory system models and context-specific joint product allocation. PhD thesis, 1998, ETH Zurich, Zurich.

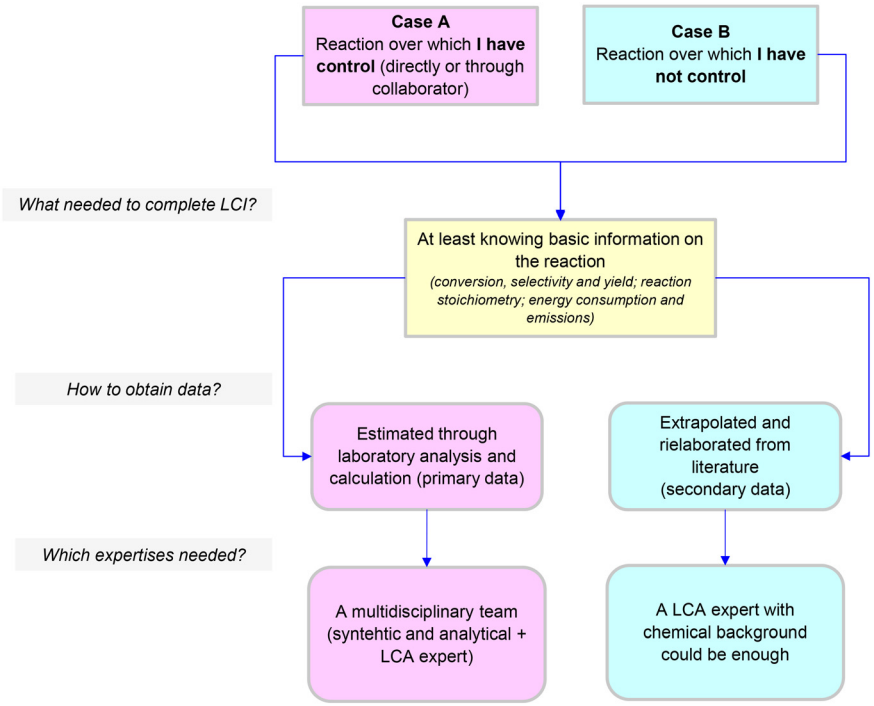


Fig. 1 Framework describing the two limit cases. (A) Reaction under control at laboratory scale, easy to obtain primary data. (B) Reaction not controlled: none direct information available; secondary data necessary to complete LCI.

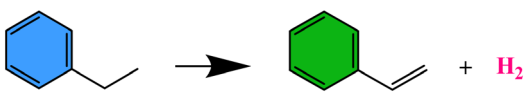
from peer-review. Ullmann's Encyclopedia of Industrial Chemistry[29] is an example of a worldwide reference document, which consultation is also suggested by literature.[21,30]

Utilizing stoichiometric reactions to gather information and complete the mass balance is advisable when the reaction consistently yields the desired product in quantitative amounts, such as the synthesis of styrene through the dehydrogenation of ethylbenzene (Scheme 1), example used by Parvatker and Eckelman.[26] In this case, conversion (C) and selectivity (S) to the desired product (styrene) are equal to 100%. As a consequence, the yield (Y) is maximum (eqn (1)).

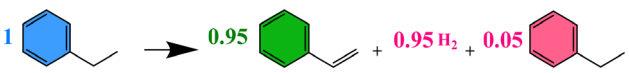
$$Y = C \times S \quad (1)$$

However, very often the reagents conversion is ≠100% (unreacted substances in the outlet flow) and/or the reactions lead to a multioutput ($S \neq 100\%$). In these cases, some alternatives are possible among the most affordable from a technical and economic point of view. The first is searching for reaction efficiency values in patents and peer-review literature. To complete the LCI in a multi-output system, it is necessary to have the selectivity values for each co-product. These values are required to calculate the yield of each output molecule based on the conversion. Patents allow a good guess of the basic information on the reaction.[30] The catalogue **https://it.espacenet.com/** can be adopted to refine the search. It is preferable to focus on international patents published within the last five years. Additionally, when examining multiple examples provided within each patent, it is recommended to choose the most suitable one, typically the example with higher C and S values relative to the desired product, as it closely resembles real-world scenarios. Nevertheless, patents have not an easy interpretation, in particular for LCA specialists with none deep experience in chemistry. In this situation, the procedure suggested by Hischier *et al.*[21] is recommended. It encourages to set the C = 95% and the S = 100% (Y = 95%). The Scheme 1 now becomes Scheme 2.

Emissions and waste streams are an integral part of the mass balance. Therefore, it is necessary to clearly define whether the system is able to recover reaction co-products or unreacted reagents and at which cost (*e.g.*, energy, auxiliary). As the approach primarily concerns laboratory or pilot scale, establishing a recovery apparatus can be challenging and



Scheme 1 Production of ethylbenzene from styrene with C = 100% and the S = 100% (Y = 100%). Blue = reagent, green = target product; pink = co-products/waste.



Scheme 2 Production of ethylbenzene from styrene with C = 95% and the S = 100% (Y = 95%), in accordance with Hischier *et al.*[21] Blue = reagent, green = target product; pink = co-products/waste.

sometimes impractical. Hence, as a broad rule of thumb, especially aligning with laboratory scale practices (although less so with pilot scale), it can be assumed that molecules not targeted for production are considered emissions. In the case of Scheme 1 the system coproduces 1 mole of H_2 (air emission) per mole of reactant, meanwhile for the reaction in Scheme 2 per each mole of reagent the system releases 0.95 moles of H_2 + 0.05 moles of C_8H_{10} (unreacted). The nature and conditions of the reaction play a significant role in determining the release compartment, which could be the atmosphere, water, or soil in the event of leakage or waste landfilled.

The procedure described above is not applicable when the stoichiometry and nature of the output molecules are not known. This situation is frequent in patents/literature articles where further analyses on mixtures not including the main product were not performed/reported by the authors. The scientific literature frequently uses the overarching term "others". They can include side products, unreacted reagents or both. If possible, this situation should be avoided. Conversely, if additional assistance is not available, one can resort to adopting the procedure outlined by Roland Hischier *et al.*,[21] described below.

Emissions into air equal to 0.2% (by mass) of the "*i*" reagent (R_i), according with Hischier *et al.*[21]

$$\text{Air emissions (AE)} = \sum_{i=1}^{n} 0.2\% \times R_i \quad (2)$$

Emissions to water, equal to the difference between the unreacted reagents and emissions to air. C_{R_i} corresponds to the conversion of R_i, according with Hischier *et al.*[21]

$$\text{Water emissions (WE)} = (100 - C_{R_i}\%) \times R_i - \text{AE} \quad (3)$$

The approach has some limitations. First, if another reagent (R_j) is involved in the reaction but only C_{R_i} of the limiting reagent is known is neither than easy complete the mass balance. Second, there are not always emissions in both environmental compartments. In these cases, an extended knowledge of the reaction is recommended. Third, Hischier *et al.*[21] did not explain the technical reasoning behind these assumptions. As the authors themselves stated, the proposed approach is merely a rough estimation and should be limited to situations where no additional data is available. Otherwise, it should be avoided.

In accordance with literature[31] the LCA model could also take into account any other fugitive emissions. If considered, they are part of the mass balance and could be estimated as follows:

- Liquid reagent with boiling point (at p = 1 atm) between 20° and 60 °C equal to 2% of the input quantity.
- Liquid reagent with boiling point (at p = 1 atm) between 60° and 120 °C equal to 1% of the input quantity.
- For a gas equal to 0.5% of the incoming quantity.

Also, in this case the physical–chemical nature of the molecules and the reaction condition play a pivotal role to determine the volatility of substances. However, it could be particularly useful in case of organic solvents.

In addition, other online platforms can be used to support the data gap. First THE MERCK INDEX§ online,[32] a web catalogue source of key physical, pharmacological and historical information on chemicals, drugs and biologicals. Since 2013, it has been updated by the Royal Society of Chemistry. Second the Reaxys® database[33] which contains 279 million substances and 65 million reactions. Both approaches are highly valuable, particularly in reconstructing retrosynthesis, especially when the building blocks are not included in the reference databases utilized for the LCA analysis.

Catalyst

The simulation of the catalytic system through LCA is everything than easy since a lot of information are covered by the corporate knowhow. First of all, the quantity of the catalyst used for the target molecule. It can differ depending on the scale. Sometimes in lab experiments an excess is used to test its efficacy and the reaction pathway. A possibility is to scale it linearly from the laboratorial to the pilot reactor, even if this could represent an overestimation. Rare information on the amount can be collected from literature. Sometimes patents, research papers, encyclopaedia chapters and best available techniques (BAT) reference documents (so called BREFs)[34] contain such communication. Otherwise, there are some alternatives. The first is asking for an expert from the academia or research centre. Her/his judgement may help in defining the amount used for the reaction at a scale of lab or proof of concept as a result of some still ongoing or just concluded research activities with open access results. Then, as reported above, a linear upscaling could be carried out. The second option is to make an estimation based on the available reaction quantity and similar reactions[35] or assume an average amount proportional to the inlet reagent or the target molecule. The latter should be coherent with the values reported in literature for similar reactions/substances. After having found the amount of catalyst used per functional unit the quantity of elements (in general metals) that constitute the active phase can be calculated by following the procedure already published in literature.[36] It consists in a stoichiometric evaluation of the quantity of each molecule or primary element involved in the final formula. Depending on the database availability the LCA software can ascribe the impacts in terms of input from nature (primary resources flows extracted) or input from technosphere (also including all the anthropogenic transformation occurred). The same procedure can be applied to the support phase. Before proceeding with the life cycle impact assessment (LCIA) stage, it's essential to verify that the analysis method incorporates characterization factors for all the metals under investigation, as already done in a previous work.[37]

§ The name THE MERCK INDEX is owned by Merck & Co., Inc., Rahway, NJ, USA and its affiliates, and is licensed to The Royal Society of Chemistry for use in the USA and Canada.

If both aforementioned strategies are not feasible, the next available option is to exclude the catalytic system from the LCA analysis. This decision can be justified if the impacts associated with the catalytic system are deemed negligible compared to other aspects of the study. This scenario may be particularly applicable when using heterogeneous catalysts, as they can be recycled for multiple cycles, potentially reducing their overall impacts over their service life. However, the exclusion of the catalytic system should be carefully considered, especially if toxic or critical raw materials are involved. For instance, excluding platinum group elements is not feasible due to their economic significance. Practitioners must at least include cradle-to-gate processes that describe ore extraction and metal purification from the technosphere (see the paragraph below on catalyst recyclability). In other cases, such as when cheaper metal oxides like Fe_2O_3 are used, exclusion is more reasonable, although it should still be avoided if possible. Therefore, any exclusions should be justified within the study (*e.g.*, lower cost and/or cut-off criteria <1% in mass). In the case of homogeneous catalysis, omitting the catalytic system might not be viable due to the complexity of the process, involving multiple synthetic steps (higher process mass intensity in organometallic complexes),[5] impactful precursors, and the inability to be recycled for multiple cycles. A recent study by Piazzi *et al.*[38] investigated the overall environmental burdens of an organometallic ruthenium-based catalyst, demonstrating its significant contribution to various impact categories within a homogeneous system. The study also showed a reduction in impacts when increasing the number of usages from 3 to 5 cycles. In this case, simulating the catalyst's environmental footprint from its synthesis was made possible through a thorough understanding of each step involved in the synthesis process at a proof-of-concept scale.

When information are not available average recycling rates can be used as a proxy:[35]

- 50% for heterogeneous catalysts.
- 0% for homogeneous catalysts.
- 90% silica gel catalysts.
- 99% for pure platinum group metals.

Energy balance

In general, accurately estimating the energy consumption of a reaction that is not under our control is far from straightforward. Therefore, two major strategies exist to overcome the problem. The first approach involves using a proxy, which can be done by either (i) employing the energy consumptions associated with the production of a similar molecule, considering factors such as chemical structure (molecular weight) and functional groups, or (ii) utilizing average data from the literature. In the first case, the values can be extracted, recalculated or deduced by already published works. Recent manuscript from Parvatker and Eckelman[20] proposed an engineering software simulation of 151 organic substances among those mostly used in reaction synthesis. A previous work from Kim and Overcash[39] shows gate to gate energy demand values for 86 chemicals, by finding those average values per kg of product: −0.5 ÷ 50 MJ (range for organic) and −4 ÷ 40 MJ (range for inorganic).

The second approach consists in using an average consumption derived from industrial measures. An example is the average electricity value for utilities equal to 1.2 MJ kg^{-1} in accordance with Althaus *et al.*,[40] also suggested by Hischier *et al.*,[21] or the 1.8 MJ kg^{-1} of thermal energy described by Huber *et al.*[35] The value from Althaus resulted from a large chemical park in Gendorf, Germany. The approach of proxy avoids to neglect the environmental impacts of the energy consumption, with a loss in accuracy. Hence, whenever feasible, it is not advisable to use proxies for estimating energy consumption. Alternatively, the methodology proposed by Andraos[41] can be applied to estimate the minimum energy required (MER) by the system. MER can be calculated by evaluating the molar enthalpy variation (kJ mol^{-1}) following variations in temperature and pressure in the reaction mixture.

The simplified equation is reported below (eqn (4)).

$$q_{tot} = q_{temp} + q_{press} = \int_{T1}^{T2} C_p(T)\mathrm{d}T + \int_{p1}^{p2} \left[V - T\left(\frac{\delta V}{\delta T}\right)_p \right] \mathrm{d}p \quad (4)$$

It represents the sum of the variations in heat (q) due to changes in temperatures ($T1$ and $T2$) and pressure ($p1$ and $p2$) that the reactants and products undergo.

For completing the energy balance, it is necessary to consider if the reaction is exothermic ($\Delta rH° < 0$) or endothermic ($\Delta rH° > 0$) by solving the eqn (5) (general formula for the enthalpy of reaction). In both cases, the net energy input (nEI) is equal to the algebraic sum between the q_{tot} and the enthalpy of reaction value, multiplied for a heat recovery efficiency value ($\eta_{heat\ rec.}$). Equation is reported below eqn (6).

$$\Delta rH° = \sum \nu \Delta fH°p - \sum \nu \Delta fH°r \quad (5)$$

$$nEI = \sum q_{tot} + \eta_{heat\ rec.} \cdot \Delta rH° \quad (6)$$

The entire procedure was applied recently to the bio-based sector.[42]

When possible, the use of MER should always be preferred by the analyst, as it provides a better understanding of what is occurring in the reaction, even though it may underestimate the impacts since it does not include the energy requirements for purification stages. As a general rule, if the reaction is at the lab scale, grid-based electricity can be used for heating. However, if simulating pilot production, major industrial energy sources such as natural gas, steam, and compressed air can be utilized.

Solvent recovery or incineration

The procedures for solvent recovery are, in general, expensive due to the energy requirements for completing distillation. In case the power (W) of each equipment is known or can be extrapolated, as well as the time of its usage (h), the simple procedure described by Gimenez *et al.*[43] can be adopted to obtain the watts consumed per hour (W h). This often results

in an overestimation of energy consumption since the instrument typically does not operate at a constant maximum power. Additionally, energy requirements may sometimes be allocated by mass or volume relative to other co-products. A similar approach was proposed for the first time by Rossi *et al.*[44] In case none information regarding the distillation procedure are available, already published data can be adopted as proxy. Jiménez-González *et al.*[45] have calculated the energy consumption for heating and cooling referred to thirteen major organic solvents (Table S1†). Those were not anymore updated, and further integrated with new greener solvents.

The LCA model may also include the step of solvent incineration. This procedure is outside the laboratory boundaries, but can occur in a pilot plant. The simplified methodology proposed is able to estimate the carbon dioxide emissions and the potential energy recovery from the incineration. The approach presented by Jiménez-González *et al.*[45] allows the usage of simplified equations. In a case of incineration without energy recovery of a spent solvent, the following values can be used to fulfil the eqn (7) (general formula for calculate the CO_2 deriving from solvent incineration):

- Combustion efficiency (η) = 99.99%.
- Combustion temperature = 900 °C.

Further efficiency values can be adopted for running a sensitivity analysis.

Carbon dioxide emissions resulting from the spent solvent incineration can be calculated as follows:

$$\text{CO}_2 \text{ emissions (kg)} = m_\text{c} \cdot \left(\frac{44\,\text{kg CO}_2}{12\,\text{kg C}}\right) \cdot \eta \qquad (7)$$

where m_C represents the carbon mass within the solvent.

The equation does not take into account other by-products resulting from the combustion process, like the NO_x and particular matter. Therefore, in case an emissions profile is available from literature (database or peer review journals) the inventory can be improved.

The unburned solvent is assumed to be released into the atmosphere in the form of VOC (volatile organic carbon) according to the following expression (eqn (8)):

$$\text{VOC emissions (kg)} = m_{\text{solvent in}}\,(1-\eta) \qquad (8)$$

where $m_{\text{solvent in}}$ represents the mass of the inlet solvent into the incinerator.

In the case the boundaries are extended to accommodate the avoided impacts deriving from the system, the following equations shall be adopted to estimate the heat flow recovered from the combustion. Eqn (9) shows the way to calculate the heat flow generated from the solvent incineration.

$$Q(\text{MJ kg}^{-1}) = \Delta H_\text{C}(900\,^\circ\text{C}) + \sum_{\text{reagents}} \nu_i \text{Cp}_i \Delta T_i + \Delta H_{\nu,\text{reagents}} \qquad (9)$$

where: ΔH_C (900 °C) = heat of combustion at 900 °C (MJ kg^{-1}), ν_i = stoichiometric coefficient of the reagent i, Cp_i = heat capacity of the reagent i (MJ kg^{-1} C°), ΔT_i = temperature difference (C°), $\Delta H_{\nu,\text{reagents}}$ = latent heat of vaporization of the reagents.

If an average recovery efficiency of 75% is assumed, the total energy recovered is equal to (eqn (10)):

$$E(\text{MJ}) = 0.75 \cdot \eta \cdot m_{\text{solvent in}}\ (\text{kg}) \cdot Q(\text{MJ kg}^{-1}) \qquad (10)$$

Further sensitivity analyses can be carried out by assuming different values of efficiency. The same approach can also be extended to unconverted reagents and emissions.

Wastewater treatment plant

Reactions, as well known, may produce some liquid effluents that should be treated into a wastewater treatment plant (WWTP), before being discharged into basin/river. Side streams, unreacted reagents (not recovered) or by-products may constitute the major contaminants. Even if the WWTP can be located outside the production boundaries, in particular for a laboratory scale reaction, practitioners may decide to include the impacts of this stage by using some proxy approach. Jiménez-González *et al.*[45] proposed a general and flexible framework that can be applied to syntheses. For organic effluents a biological WWTP can be assumed, by estimating the main parameters on the basis of the COD (cumulative oxygen demand) entering the system. In the publication, they provide general data for an average biological WWTP able to treat 1 kg of COD.

Two general equations can be settled, in order to describe the main mechanism. The first describe the cell auto-oxidation (Scheme 3):

Substrate represents the organic chemicals generated from the synthesis and which enters the WWTP (S_0). Jiménez-González *et al.*[45] proposed $C_6H_{12}O_6$ as a model molecule for it. After a certain time it is reduced (S_r). a represents the kg of cells produced per kg of volatile suspended solids of COD reduced. a' is the kg of oxygen consumed per kg of inlet COD. b the kg of cell undergoing autoxidation per kg of total cells and b' is the kg of oxygen consumed per kg of total cells. $C_5H_7NO_2$ (molecular weight 113.11 g mol^{-1}) can be used as a model molecule for cells. The amount of oxygen necessary for the process can be estimated as follow:

$$\begin{aligned} \text{O}_2 &= a'S_0 + b'X_\text{ML}\ (1) \\ X_\text{p} &= aS_\text{r} - bX_\text{ML}\ (2) \end{aligned} \qquad (11)$$

X_ML represents the mass of total cells in the mixture (estimated also as the ratio between food and microorganisms) and X_p the net biosolid production (sludge then sent to stabilization and landfill). Once known the mount of O_2 the

Substrate + $a'O_2$ + NH_3 → a microbial cells + $a'CO_2$ + $4H_2O$ (1)

b microbial cells + $b'O_2$ → $b'CO_2$ + H_2O + NH_3 (2)

Scheme 3 General formula for cell auto-oxidation, according with Jiménez-González *et al.*[45]

amount of CO_2 can be estimated by the following relationship based on molecular weights ratio (44/32 = 1.375):

$$CO_2 = 1.375O_2 \quad (12)$$

Often only b values are tabulated, and b' can be recalculated using the following formula

$$b' = 1.42b \quad (13)$$

where 1.42 is a constant deriving from the ratio between the molecular weight of cells and oxygen based on the following equation (Scheme 4):

According to literature[45] the average efficiency for COD removal is 86%. In general, two major energetic vector are used in WWTP: electricity and gasoline. The first for aeration (47%), anaerobic treatment of biosolids (19%), pumping (13%), others miscellaneous (21%). Diesel is used for vehicles. Ammonia and orthophosphate are in general added as nutrients, by respecting the relationship 100(BOD):5(N):1(P). Polymers and NaOH are also used as further ancillary substances. Sodium hydroxide is then discharged in the effluent. For inorganic salts free from heavy metals, 90% will remain in the effluent and the rest absorbed in the biosolids. In the case they contain heavy metals only the 25% remain in the effluent. Jiménez-González *et al.*[45] collect average value for simulating a WWTP, summarized in Table 1.

Predictive methods

Predictive methods involve using regression models to predict the environmental impacts of molecules based on easily obtainable descriptors, such as molecular weight, number, and type of functional groups. These methods are typically not used to fill inventory data (even if done in the past)[46] but rather to obtain results in terms of predefined environmental indicators. Several examples exist in the literature.

The first predictive model based on artificial neural networks (ANNs) was FineChem,[47,48] which was developed based on pioneering work comparing the prediction accuracy of ANNs with linear regression.[49] Subsequent modifications aimed to enhance predictive ability, such as data processing strategies and the use of trained multilayer ANNs.[50,51]

Researchers have demonstrated that the performance of predictive models can be further improved by incorporating additional thermodynamic descriptors.[52,53] Linear and non-linear regression models have also been combined for predictive LCA results.[54]

These predictive methods have been applied in various fields, including green chemistry, to compare alternative syntheses,[55] evaluate the benefits of molecule substitution,[56] and perform environmental and economic assessments of biorefineries.[57]

Scheme 4 General formula for cells oxidation, according with Jiménez-González *et al.*[45]

Table 1 Average LCI for WWTP

	COD basis	TOC basis
COD_{in}, kg	1.00×10	—
TOC_{in}, kg	—	1.00×10
$Inorganic_{in}$, kg	6.60×10^{-1} (0.06 heavy metal salts)	1.90×10 (0.2 heavy metal salts)
Electricity for WWTP, kW h	1.10×10	3.10×10
Ancillary substances (polymer, ammonia, NaOH, phosporic acid)	6.00×10^{-2}	1.80×10^{-1}
Fuel, kg	5.00×10^{-3}	1.00×10^{-2}
COD_{out}, kg	1.40×10^{-1}	—
TOC_{out}, kg	—	1.40×10^{-1}
$Inorganic_{in}$, kg	5.55×10^{-1} (0.015 heavy metal salts)	1.58×10 (0.05 heavy metal salts)
$NaOH_{out}$, kg	1.30×10^{-2}	3.80×10^{-2}
CO_2 emission, kg	8.70×10^{-1}	2.49×10
Biosolids organic, kg (dry)	3.50×10^{-1}	1.00×10
Biosolids ancillary, kg (dry)	4.70×10^{-2}	1.42×10^{-1}

Recently was launched FineChem 2, which aimed to improve the prediction of carbon footprint values of chemicals by combining machine learning with first-hand industry data.[58]

While these methods may not be designed for completing inventories, practitioners can use them to complete the inventories and extrapolate key environmental results for molecules where traditional inventory compilation methods are not applicable. The results obtained, normalized for the amount used by the system, can be added to other LCA results to provide a cumulative snapshot. Although these methods may not be suitable for conducting hotspot analyses,[59] they can help cover compounds falling within their scope that have not yet been included in LCIs.

Exclusion and comparison with previous approached

Some other aspects are not discussed here since out of the scope of this manuscript for several reasons. Among those, the following are noteworthy.

Second the use of the Process Economics Program (PEP) Yearbook[60] to complete the mass and energy balances. PEP reports represent one of the more consolidated data sources for chemicals, since collect in yearbook primary data extrapolated from companies and associations, concerning raw material and energy consumptions, catalyst amount and price, forecast and outlook. However, as already debated in literature[30] subscription is rather expensive. Therefore, they cannot be included as reference documents to fill data gap in low budget studies.

In addition, the inclusion if the chemical infrastructure (*e.g.*, plant and part thereof) in the LCI. They were not discussed before since negligible in terms of environmental impacts, due to the high-rate usage. However, proxy data exist[21] and they can be used to compute the impact of the infrastructure and thus checking the results variability (sensitivity analysis).

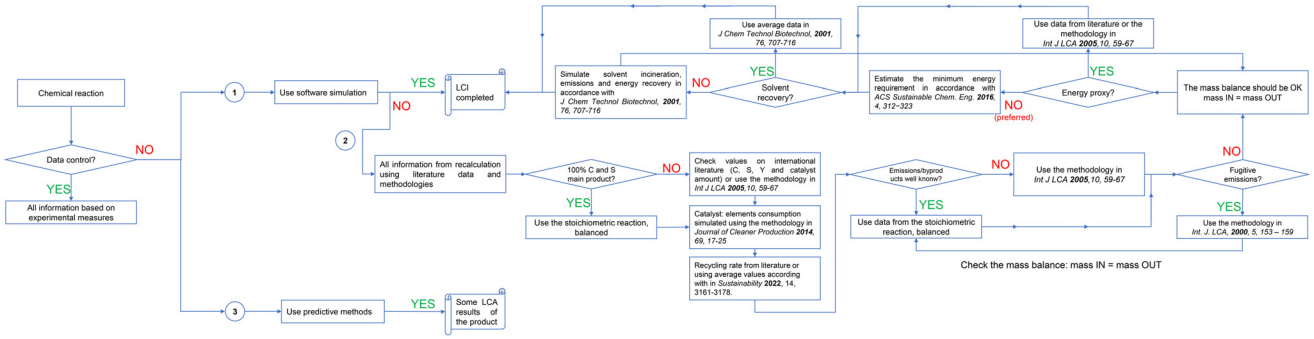


Fig. 2 Procedural approach to fill data gap in LCA of chemicals: a simplified scheme.

Fig. 2 shows the whole procedural approach presented above to fill the data gap in LCA of chemicals.

The flowchart proposed, as well as the content of whole manuscript get inspiration from previous studies proposed for addressing data gaps (all cited above). Among those, one first attempt is represented by Jiménez-González *et al.*,[31] who proposed a methodology to support practitioners to fill gate-to-gate LCA studies at industrial scale (functional unit proposed 1000 kg). We are at the begin of 2000s when the methodology started to take the lead among the tools to assess the environmental loads, even if not yet standardized. They guide the readers among the selection of the more consolidated industrial synthetic process, and its definition through the mass balance by also including fugitive emissions. Material Safety Data Sheets can be used to understand the presence of impurities. They suggest to express the energy consumption in MJ, and to take into consideration: the heat of reaction, the heat of dilution, the energy for materials transportation and distillations, *etc.* However, a strong chemical background and knowledge of the reaction under study is required to complete all the steps. In addition, the work does not cover the waste treatment modules, discussed later by some of the authors[45] who also proposed an application in the oil refining sector.[61] Some years afterwards, Hischier *et al.*[21] pointed out the limited population of LCA databases. They guide the readers through a procedural approach to fill the data gap, developed within their activity of populating the ecoinvent database. Some general instructions are given, among those the necessity of using average mass efficiency of 95% and energy consumption derived from an industrial plant in Germany (Gendorf). They also suggest the usage of Ullmann's Encyclopedia of Industrial Chemistry to retrieve information, as done previously by Bretz and P. Frankhauser[30] two pioneers in this filed given their work at CIBA-Geigy Limited. In 2018 Yao and Masanet proposed a simplified approach, so-called pathways assessment modelling framework (PAM), to evaluate the energy consumption and the emissions of emerging technologies in the chemical industry.[62] The approach is very useful when the reaction is under control (direct or indirect), in order to have complete mass and energy balances. This allows the users to fulfil the method by retrieving embodied impacts from databases. In line with the manuscript here proposed, and with Andraos,[41] they encourage the estimation of energetic consumption from the unit operation module starting from thermodynamics. The approach suggests the usage of Monte Carlo for evaluating the uncertainties, due to the lower TRL (3–5). Parvatker and Eckelman[26] compare different approaches for filling data gap in LCI. To my knowledge, this represents the first example of a ranking between the different pathways examined, in terms of accuracy and time consumption. Getting primary data and software engineering simulation represents the top level. However, as described above, they require a deep knowledge of the reaction and cannot be applied when there is no control on the synthesis. The procedural approach here ranges from method 2 and 6 of the scale proposed by Parvatker and Eckelman, since thought to guide the user in the right choice (that could be different case by cases) in order to avoid the step 7: omit flows. More recently, Huber *et al.*[35] proposed the RREM approach (research, reaction, energy, and modelling) based on first research on the chemical and its synthesis process. Followed by a setup of the reaction equations and a cross-checking with existing databases, then on the thermal energy demand. Finally, to model the dataset and connect it to existing ones. The method is very useful, even if some aspects are missing such as the catalyst simulation, the solvent/unreacted reagent incineration, the waste water treatment and the fugitive emissions. Hence, in the compendium proposed the idea is to collect all the efforts already published and propose a flowchart, which may guide the user to complete the LCA study. It is not intended as an alternative to previous works. The LCA practitioner may decide to fully adopt it or to adapt it to the case study.

Results and discussion

In the previous section, several methodological approaches were presented to fill the data gaps for chemical reactions whose inventories are not under control of the LCA practitioner. Regarding the energy issue, the dataset provided by Parvatker and Eckelman[20] for the 151 organic substances can serve as a basis for comparing the results obtained with the

number of H and C atoms, even if somehow limited since no data were available in the range C12 ÷ C14. The trend obtained is illustrated in the Fig. 3.

In all instances, variability exists around the average value represented by the black diamond. However, utilizing these means can provide assurance in covering energy consumption, thus avoiding costs for additional analyses or simulations, which may require more resources or prolong the duration of the study. Understanding the average energy consumption relative to the number of carbon and hydrogen atoms can facilitate the adoption of these values even when dealing with unknown molecules but with some defined information regarding their chemical structure, such as in the case of certain patents. Additionally, knowledge of the minimum and maximum values can be valuable for conducting sensitivity analyses to demonstrate how the final impact scores are influenced by this variable. The usage of such type of investigation is strongly recommended in prospective analyses.[63]

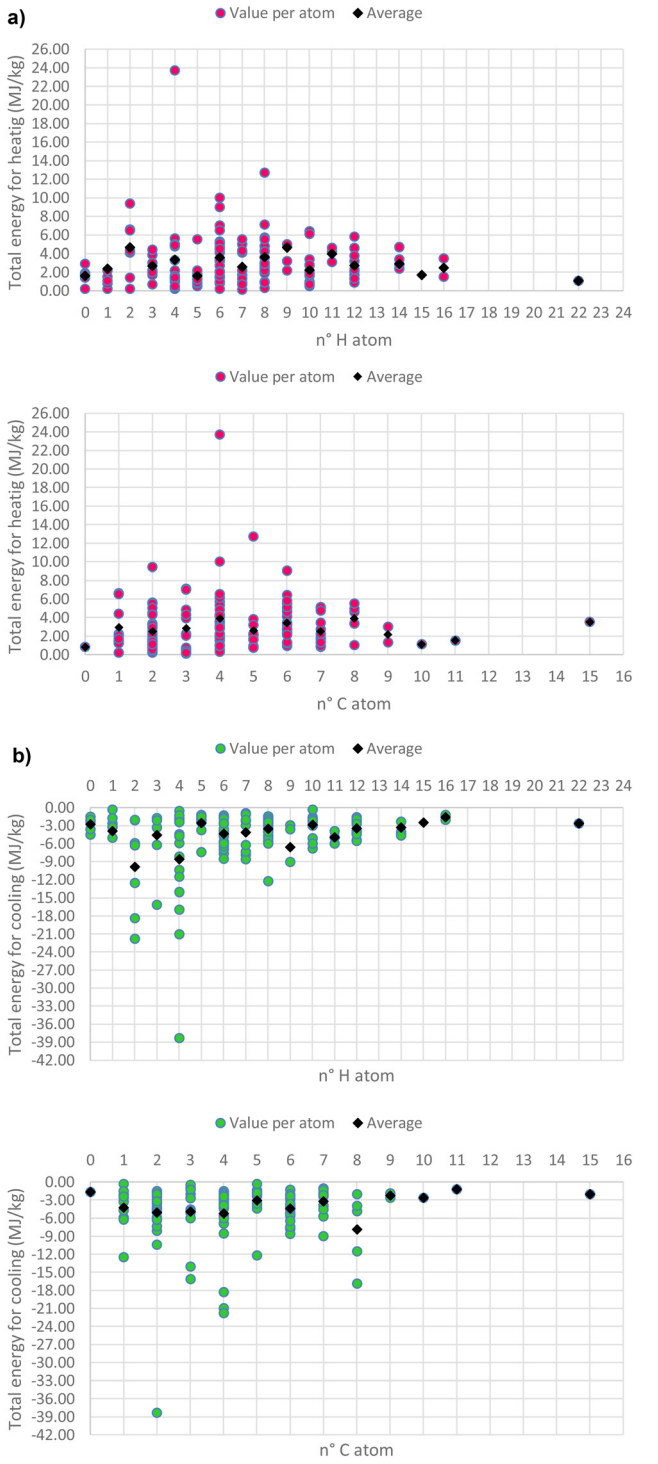


Fig. 3 Energy consumption in MJ kg^{-1} for heating (a) and cooling (b) procedures plotted per the number of C and H atoms. Data were extrapolated from literature.[20]

All results of the interpolation are reported in ESI (Tables S2 and S3; Fig. S1–S2†). The average energy consumption for heating and cooling appears to increase and decrease with the number of carbon atoms, though this correlation is not easily explained. Generally, heat requirements are influenced by the nature of the reaction (exothermic or endothermic) as well as the energy required for cooling. For 95% of the dataset provided by Parvatker and Eckelman,[20] the reactions occur above room temperature, with 25 reactions occurring at temperatures above 300 °C. This fluctuation could be associated with these characteristics and may also be influenced by the references used to create the software simulation for each chemical. A further correlation between energy requirements and molecular mass (MM) was also examined (Fig. S3 and S4†). The sample was divided into three subsets based on molecular weight: (a) 30 < MM < 69, (b) 70 < MM < 136, and (c) MM > 137. By plotting MM against energy consumption (for heating and cooling), the values fall within specific ranges.

- For subset (a), the energy for heating falls within the range of 0.2 to 10.0 MJ kg^{-1} for 31 out of 32 molecules, and for cooling, within the range of −0.7 to −9.7 MJ kg^{-1}.
- For subset (b), 96 out of 97 molecules fall within the heating range of 0.1 to 8.6 MJ kg^{-1}, and 95 out of 97 fall within the cooling range of −0.3 to −9.6 MJ kg^{-1}.
- For subset (c), all molecules fall within the heating range of 0.8 to 5.0 MJ kg^{-1} and the cooling range of −1.1 to −4.7 MJ kg^{-1}.

Those molecules that appear as outliers need further investigation before being excluded from the dataset, as in the case of furan, methyl methacrylate and acrylic acid (for cooling).

Even though it is easier to apply, interpolating energy consumption based on the number of hydrogen and carbon atoms introduces additional uncertainties, as it may not consider all relevant factors affecting energy consumption. For this reason, where possible, other alternatives should be applied.

The approach proposed by Andraos[41] presents some difficulties and limitations. For example, the success of the enthalpy balance is linked to a good knowledge of the reaction involved, of basic concepts of thermodynamics and of mathematical analysis (not always so obvious). In addition, it does not take into account all the energy consumption, but represents a minimum requirement of the system for the reaction

to take place. For example, it does not include the energy requirements of the separation and regeneration procedures for the catalyst (in particular if heterogeneous), or the energy consumption for the distillation steps to recover solvents.

Despite these aspects, if completed it returns a plausible impact compared to consumption on a laboratory and pilot scale and is easily repeatable for other reactions.

Case study

To test the feasibility of the approach proposed above, exploring the variability in results, the case study of maleic anhydride (MA) produced at pilot scale from bio-butanol (from dedicated crops, 1/3 maize, 1/3 sugarcane and 1/3 switchgrass) is presented. The geographical boundaries are settled in Europe. The synthesis, full inventory and LCIA results were largely discussed in a recent publication.[42] All models were created using SimaPro[64] software (v.9.6.0.1). Ecoinvent database[65] (v.3.10) was used for simulating all the background information, by selecting the market scenarios (to include impacts from average transportation distances) and the allocation at the point of substitution (APOS) unit models. In this case, the method developed by the Intergovernmental Panel of Climate Change (IPCC 2021, GWP100 incl. CO_2 uptake, v.1.01)[66] and the and the CED (Cumulative Energy Demand, v.1.11)[67] were used as key environmental indicators to complete the cradle-to-gate simulation. Both represent single-impact methods, as they express results focusing solely on one environmental issue: the global warming potential (*i.e.*, carbon footprint) and the consumption of renewable and non-renewable resources (*e.g.*, fossil and bio-based raw materials, fuel, *etc.*). Single-issue methods are generally effective for communication, even though a LCA requires a comprehensive environmental profile that includes multiple burdens. These two methods were chosen here because they are easier to convey to a broader audience, including synthetic chemists and LCA practitioners. A more comprehensive impact assessment of bio-based maleic anhydride was conducted in the previous work.[42]

Three main scenarios were considered in terms of mass efficiency (mass only, MO):

(1) C = 100% and S = 100% (Y = 100%), in which MA is co-produced together with water then sent to waste water treatment (WWT) plant (MO:C1S1Y1 scenario) (Scheme 5). The reaction can be summarized according to following equation.

(2) C = 95% and S = 100% (Y = 95%), according with Hischier *et al.*[21] the output flow is characterized by an unreacted amount of the butanol (5% of the inlet moles) then sent to incinerator (Scheme 6). Air emissions in terms of biogenic CO_2 (from butanol combustion) and volatile organic compounds (VOCs – from the unconverted fraction) were calculated from Jiménez-González *et al.*[45] Water is sent to WWT (MO:C1S1Y0.95).

(3) C = 100% and S = 50% (Y = 50%), in which MA is co-produced together with water and carbon dioxide (MO:C1S1Y0.5) (Scheme 7). The first is assumed to be sent to WWT, the latter to be emitted in the atmosphere as biogenic CO_2.

Full mass balances are reported in ESI (Tables S4–S6†). The catalyst was omitted, since its simulation was already discussed in the previous publication.[42]

Regarding the energy flows the model was compiled according with:

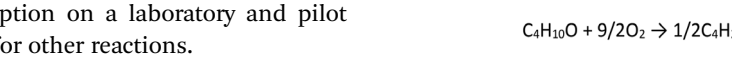
- Andraos (A),[41] to estimate the MER of the reaction. Three scenarios were created (4) MA:C1S1Y1, (5) MA:C1S1Y0.95 and (6) MA:C1S1Y0.5. Full energy balances are reported in ESI (Tables S7–S9†). According to literature,[42] nEI values since negative (benefit for the system) were assumed in the form of process steam recovery from the whole system (*i.e.*, avoided steam production from traditional sources);
- Parvatker and Eckelman,[20] by creating six scenarios. The first three consider to use the proxy energy values (for cooling and heating, P = proxy) reported for the MA produced from benzene and butane, respectively (7) MP:C1S1Y1, (8) MP:C1S1Y0.95 and (9) MP:C1S1Y0.5. The latter were built using the average energy values (for cooling and heating) obtained for a C_4 molecule (PC), Table S3,† respectively (10) MPC:C1S1Y1, (11) MPC:C1S1Y0.95 and (12) MPC:C1S1Y0.5.

Fig. 4 and 5 collect the comparison among the twelve scenarios investigated, by assessing the production of 1 kg of MA (functional unit or FU) in terms of carbon footprint and CED indicators respectively. The graphs show the relative contribution of the sub-categories,¶ and the net values (blue rhombus). A Monte Carlo simulation was carried out to evaluate the effect of the uncertainties on the cumulative scores. The lognormal statistical distribution, with a 95% confidence interval and an iterative calculation number of 1000 simulations, was applied.

As depicted in the figures a variability exists on the final scores, both in terms of absolute values (net) and relative contribution of each subcategory (please also check Tables S10 and S11†). However, as we move from left to right, the overall

Scheme 5 Stoichiometric reaction, molar values respect to MA: C = 100% and S = 100% (Y = 100%).

Scheme 6 Stoichiometric reaction, molar values respect to MA: C = 95% and S = 100% (Y = 95%).



Scheme 7 Stoichiometric reaction, molar values respect to MA: C = 100% and S = 50% (Y = 50%).

¶Subcategories in the case of IPCC 2021: biogenic; CO_2 uptake; fossil and land transformation. Subcategories in the case of CED: non-renewable, biomass; non-renewable, nuclear; non-renewable, fossil; renewable, biomass; renewable, water and renewable, wind, solar, geothermal.

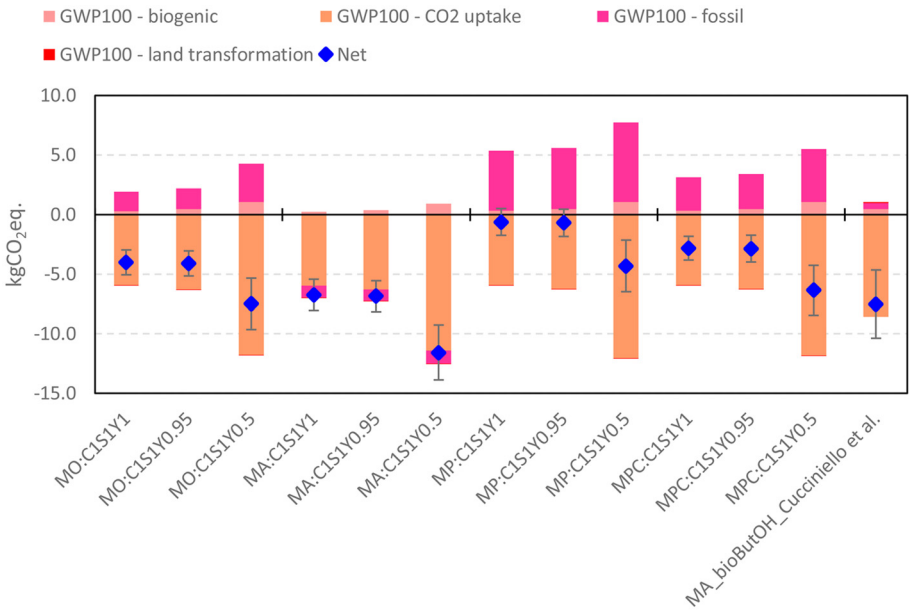


Fig. 4 MA production from bio-butanol, comparison of different scenarios using the method IPCC 2021, GWP100 incl. CO_2 uptake, v.1.01. (1) MO:C1S1Y1 *C* 100%, *S* 100%, *Y* 100%; (2) MO:C1S1Y0.95 *C* 95%, *S* 100%, *Y* 95%; (3) MO:C1S1Y0.5 *C* 100%, *S* 50%, *Y* 50%; (4) MA:C1S1Y1 [(1) + Andraos], (5) MA:C1S1Y0.95 [(2) + Andraos]; (6) MA:C1S1Y0.5 [(3) + Andraos]; (7) MP:C1S1Y1 [(1) + Proxy MA]; (8) MP:C1S1Y0.95 [(2) + Proxy MA]; (9) MP:C1S1Y0.5 [(3) + Proxy MA]; (10) MPC:C1S1Y1 [(1) + Proxy C4 avg]; (11) MPC:C1S1Y0.95 [(2) + Proxy C4 avg]; (12) MPC:C1S1Y0.5 [(3) + Proxy C4 avg] and (13) Cucciniello *et al.*[42]

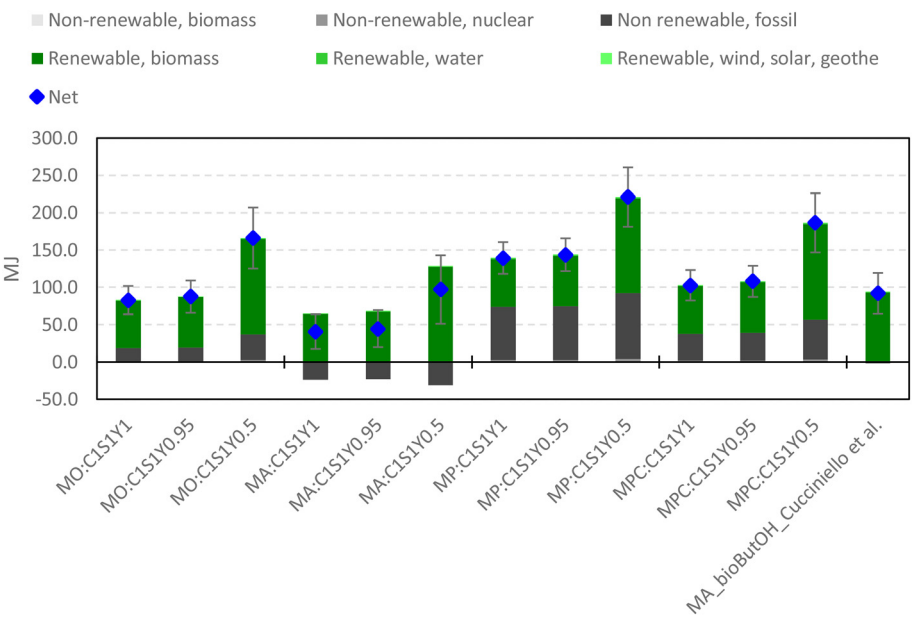


Fig. 5 MA production from bio-butanol, comparison of different scenarios using the method CED, v.1.11. (1) MO:C1S1Y1 *C* 100%, *S* 100%, *Y* 100%; (2) MO:C1S1Y0.95 *C* 95%, *S* 100%, *Y* 95%; (3) MO:C1S1Y0.5 *C* 100%, *S* 50%, *Y* 50%; (4) MA:C1S1Y1 [(1) + Andraos], (5) MA:C1S1Y0.95 [(2) + Andraos]; (6) MA:C1S1Y0.5 [(3) + Andraos]; (7) MP:C1S1Y1 [(1) + Proxy MA]; (8) MP:C1S1Y0.95 [(2) + Proxy MA]; (9) MP:C1S1Y0.5 [(3) + Proxy MA]; (10) MPC:C1S1Y1 [(1) + Proxy C4 avg]; (11) MPC:C1S1Y0.95 [(2) + Proxy C4 avg]; (12) MPC:C1S1Y0.5 [(3) + Proxy C4 avg] and (13) Cucciniello *et al.*[42]

trend depicted across the three mass efficiency scenarios remains consistent within each cluster (Mass only; Mass + Andraos; Mass + Proxy MA; Mass + Proxy C_4 avg). In the case of IPCC from MO:C1S1Y1 to MO:C1S1Y0.5 fossil and biogenic indicators rise up +99% and +240% (respect to MO:C1S1Y0.95 + 90% and + 124%) as a consequence of the lower efficiency of the reaction (from 100% to 50% molar yield) which implies more reagent per functional unit. However, a greater usage of

dedicated biomasses to produce butanol allows a higher CO_2 uptake that reduces the overall carbon footprint values. Hence, in accordance with ISO 14067,[68] the results of a carbon footprint shall be presented per each sub-indicator and not in a cumulative form. This approach helps prevent misleading interpretations and supports accurate communication of the results. Tables S12 and S13 in the ESI† present the mean, standard deviation (SD), and coefficient of variation (CV) values for IPCC and CED results. The CV, which is the ratio between the SD and the mean, indicates the variability of the final results per category. A higher CV value signifies greater variability and can be used to rank data by the relative magnitude of uncertainty. The table shows that for the IPCC method, CO_2 uptake and land transformation have the highest CV scores in all scenarios, except for those where the Andraos formula for MER was applied (scenarios MA:C1S1Y1, MA:C1S1Y0.95 and MA:C1S1Y0.5), in which fossil CV dominates over carbon sequestration. These scenarios are the only ones where the fossil index has negative values. Therefore, although carbon sequestration during growth significantly impacts the net carbon footprint (with land transformation having a negligible contribution), the indicator results are subject to the highest degree of uncertainty. It is well-known that carbon sequestration during plant growth is highly influenced by the fixation capacity, which varies over the years. CED, on the contrary, depicts a greater cumulative consumption of resources moving from MO:C1S1Y1 to MO:C1S1Y0.5 + 101% (+90% respect to MO:C1S1Y0.95), due to the increase of the non-renewable fossil and renewable biomass resources that double in both cases. The usage of MRE to fill the energy balances (MA:C1S1Y1, MA:C1S1Y0.95 and MA:C1S1Y0.5) contributes in reducing the cumulative carbon footprint value, thanks to the energy recovery in the form of steam which has beneficial effects on the fossil category. The same is shown for CED where the dark grey category assumes negative values in the range −23 ÷ −31MJ. The scenarios which use proxy data for energy (from MP:C1S1Y1 to MPC:C1S1Y0.5) represent those more conservative in terms of environmental impacts, since they achieve the higher burdens in both methods used. The variability is mainly due to the energy mix used to simulate heating and cooling activities, in this case natural gas was selected since one of the major energy vectors in chemical industry.[69] In order to reduce the overall score and the contribution of the fossil subcategories (CO_2 eq. and MJ) a greater share of renewable energy should be adopted. To better understand the potential impact of changing the energy source, Tables S14 and S15† compile the impact results of various major energy vectors to illustrate their variability. The average European electricity mix was used as a reference, along with Spanish electricity from solar thermal parabolic panels and German onshore wind production (representing major EU countries in each field). As shown, the higher the percentage of renewable energy, the lower the cumulative carbon footprint value. A similar trend is observed in terms of non-renewable sources.

Additionally, a final histogram was added to compare the results with those previously achieved by Cucciniello *et al.*[42] for the same case study. The dataset used in the previous publication is more reliable, even though no primary data were available. Efficiencies already tested at the lab scale were used to complete the mass balances, including co-products and emissions.[70]

The catalyst, as mentioned earlier, was simulated using the procedural approach previously published.[36] The energetic input was compiled through the MER. Nonetheless, the results appear comparable in terms of both carbon footprint and resource consumption, in particular with the scenarios in which thermodynamics was used to complete the energetic parts, even if in Cucciniello *et al.*[42] the lower yield (39%) allows a resources consumption which is mainly due to dedicated biomasses. The use of proxies still results in higher burdens and should be considered if a worst-case scenario needs to be analysed or if a sensitivity is necessary.

As shown the uncertainty level plays an important role in defining the right results, since values can be 2 to 5 times greater than the mean and they seem in line with findings in literature for the petrochemical sector (*e.g.*, I the case of butadiene).[71] The authors investigated the uncertainties in greenhouse gas emissions estimates for petrochemical production. They analysed cradle-to-gate emissions of 81 chemicals based on 2043 types of chemical manufacturing process at 37 379 facilities worldwide, assessing six sources of uncertainty related to: (i) the choice of the allocation method to treat co-products, (ii) the heterogeneity of processes for manufacturing the same chemical; (iii) the embedded in upstream feedstocks; (iv) the indirect energy use; (v) the direct energy use and (vi) the direct processes (chemical reactions). Some uncertainty sources more related to the system boundaries choices were not included (*e.g.*, cradle-to-grave approach, the temporal correlation, the technology readiness level of processes considered). However, the results estimate a 34% uncertainty in total global emissions of 1.9 ± 0.6 Gt CO_2eq. for 2020, with 15–40% uncertainties across most of the petrochemicals analysed. The largest uncertainties arise from the inability to assign specific production processes to facilities due to data limitations. Uncertain data on feedstock production and off-site energy generation significantly contribute. To reduce uncertainties throughout the industry, the most valuable target is therefore upstream chemicals and in particular primary chemicals, where owing to high production volumes, ethylene, propylene and ammonia have the largest absolute uncertainties. On the other hand, on-site fuel combustion and chemical reactions play smaller roles, as well as the choices of allocation methods (generally insignificant). They found that prioritizing the specification of facility-level processes in data collection for just 20% of facilities could reduce global uncertainty by 80%. This highlights the necessity of quantifying uncertainty in petrochemical greenhouse gas emissions globally and outlines priorities for improved reporting. The generated dataset offers independent emissions factor estimates based on facility-specific information for 81 chemicals, supporting future analyses. Another example is represented by Khoo *et al.*, which provide a set of 8 criteria to ensure data

quality applied for various types of LCA cases. They suggest a guideline to rank data quality to help ensure that LCA models and results are reliable enough to make sound conclusions. Therefore, it is necessary to point out the importance of uncertainty in LCA results, by also refer to the existing national and international standards on the argument.[72–74]

To conclude, the results here presented describe this particular case study and they cannot be taken as an absolute reference since valid with the boundaries considered. However, as a rule of thumb, the assessment of chemical reactions is always possible even if they are not under the direct control of the LCA practitioner. The application of data ranking criteria for chemical synthesis[25] is always recommended before proceeding with the assessment, as well as the inclusion of a sensitivity analysis to explore all the possibilities and verify the case in which the results are competitive respect to a benchmark. In addition, the LCIA should cover more impact categories beyond carbon footprint and resource consumption. While these indicators are useful, they are limited and can sometimes be misleading. Therefore, it is strongly recommended to include a broader range of impact categories. This approach reduces the likelihood of greenwashing and prevents "impact shifting", where impacts are transferred from one category to another that is not included in the assessment.

Conclusions

Having the capability to include a life cycle simulation to support the work of synthetic chemists was underlined as a key element to guide green chemistry innovation. However, sometimes the lack of data inhibits the analysis. In this critical review a step toward a procedural life cycle inventory of chemicals, at laboratory and pilot scale, was proposed to cover these gaps and overcome the unavailability when the reaction is not under the direct control of the LCA practitioner. The approach here proposed collects the main efforts already suggested in literature in the form of a procedural guideline to suggest the right choice and encompass barriers (Fig. 2). The synthesis of bio-based maleic anhydride from butanol was selected as a case study to discuss the results obtained from the application of the approaches proposed. Two single-issue methods were selected for completing the LCIA stage: the carbon footprint (IPCC 2021, GWP100 including CO_2 uptake, v.1.01) and resource consumption (CED, v.1.11). Although both are user-friendly and familiar to researchers not well-versed in life cycle assessment, a multiple-impact approach is in general recommended to support decisions, as it can cover a broader range of environmental burdens. The results show a great variability exists in the final scores, both in the case the discussion ends at the subcategories level or is extended to the net values. Mass and energy balances directly affect the final results, even if a trend is depicted shifting from scenarios with lower to greater material efficiency. The inclusion of MRE allows to rebalance the scores, due to the higher exothermic grade of the reaction and the potential steam recovery (at pilot scale). On the other hand, the usage of proxy energy values for MA (from fossil) or extrapolated for average C_4 substances guarantee more conservative footprint scores with higher LCIA results. As a general rule of thumb, a sensitivity analysis should be always included in the study to explore alternatives. In addition, the LCIA stage shall be extended to a wider range of impact categories rather than an evaluation of the carbon footprint only. When possible, a multiple-issue methodology is recommended. The method here proposed has some limitations, among those the difficulty to be applied to some class of chemicals. This could be the case of polymers or nanoparticles, when an exact stoichiometric reaction is not known. However, as suggested by literature,[30,75] those approaches remain the better solution to overcame lack of data or resources (time and money). In general, there is no a one-fits-all strategy. This is mainly due to the wide variety of chemical processes and the dataset availability at the moment of the analysis. A blended usage of the various approaches discussed before may allow to combine their strengths, by preventing gaps. In addition, the inclusion of a sensitivity analysis results to be important in order to present the final environmental result in a relative form rather than absolute.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The author would thank all the members of the research group in environmental and cultural heritage chemistry of the University of Bologna, in particular Fabrizio Passarini, all the past and present students, the colleagues and firends Raffaele Cucciniello, Fabrizio Cavani and Paul T. Anastas for their support during the years.

References

1 International Organization for Standardization (ISO) (2006a) 14040:2006, Environmental Management—Life Cycle Assessment—Principles and Framework. Geneva, Switzerland: ISO.
2 International Organization for Standardization (ISO) (2006b) 14044:2006, Environmental Management—Life Cycle Assessment—Requirements and Guidelines. Geneva, Switzerland: ISO.

3 D. Cespi, F. Passarini, E. Neri, R. Cucciniello and F. Cavani, LCA integration within sustainability metrics for chemical companies, in *Life Cycle Assessment in the Chemical Product Chain: Challenges, Methodological Approaches and Applications*, ed. S. Maranghi and C. Brondi, 2020, pp. 53–73.
4 A. D. Curzons, D. J. C. Constable and V. L. Cunningham, *Clean Prod. Process.*, 1999, **1**, 82–90.
5 A. D. Curzons, D. J. C. Constable, D. N. Mortimer and V. L. Cunningham, *Green Chem.*, 2001, **3**, 1–6.
6 P. Saling, A. Kicherer, D. Dittrich-Krämer, R. Wittlinger, W. Zombik, I. Schmidt, W. Schrott and S. Schmidt, *Int. J. Life Cycle Assess.*, 2002, **4**, 203–218.
7 A. Shonnard, A. Kicherer and P. Saling, *Environ. Sci. Technol.*, 2003, **37**, 5340–5348.
8 P. Saling, R. Maisch, M. Silvani and N. König, *Int. J. Life Cycle Assess.*, 2005, **10**, 364–371.
9 A. D. Curzons, C. Jiménez-González, A. L. Duncan, D. J. C. Constable and V. L. Cunningham, *Int. J. Life Cycle Assess.*, 2007, **12**, 272–280.
10 P. J. Dunn, S. Galvin and K. Hettenbach, *Green Chem.*, 2004, **6**, 43–48.
11 C. Jiménez-González, A. D. Curzons, D. J. C. Constable and V. L. Cunningham, *Int. J. Life Cycle Assess.*, 2004, **9**, 114–121.
12 C. Jiménez-González, C. Ollech, W. Pyrz, D. Hughes, Q. B. Broxterman and N. Bhathela, *Org. Process Res. Dev.*, 2013, **17**, 239–246.
13 L. Leseurre, C. Merea, S. Duprat de Paule and A. Pinchart, *Green Chem.*, 2014, **16**, 1139–1148.
14 T. E. Swarr, R. Cucciniello and D. Cespi, *Green Chem.*, 2019, **21**, 375–380.
15 International Organization for Standardization (ISO) (2006) 14025:2006, Environmental labels and declarations—Type III environmental declarations—Principles and procedures. Geneva, Switzerland: ISO.
16 International Organization for Standardization (ISO) (2014) 14046:2014, Environmental management—Water footprint—Principles, requirements and guidelines. Geneva, Switzerland: ISO.
17 International Organization for Standardization (ISO) (2018) 14067:2018, Greenhouse gases—Carbon footprint of products—Requirements and guidelines for quantification. Geneva, Switzerland: ISO.
18 H. C. C. Erythropel, J. Zimmerman, T. de Winter, L. Petitjean, F. Melnikov, C. H. Lam, A. Lounsbury, K. Mellor, N. Jankovia, Q. Tu, L. Pincus, M. Falinski, W. Shi, P. Coish, D. Plata and P. Anastas, *Green Chem.*, 2018, **20**, 1929–1961.
19 D. Kralisch, D. Ott and D. Gericke, *Green Chem.*, 2015, **17**, 123–145.
20 A. G. Parvatker and M. J. Eckelman, *ACS Sustainable Chem. Eng.*, 2020, **8**, 8519–8536.
21 R. Hischier, S. Hellweg, C. Capello and A. Primas, *Int. J. Life Cycle Assess.*, 2005, **10**, 59–67.
22 E. Heinzle, D. Weirich, F. Brogli, V. H. Hoffmann, G. Koller, M. A. Verduyn and K. Hungerbühler, *Ind. Eng. Chem. Res.*, 1998, **37**, 3395–3407.
23 H. H. Khoo, V. Isoni and P. N. Sharratt, *Sustain. Prod. Consum.*, 2018, **16**, 68–87.
24 U.S. Energy, Information Administration (eia), **https://www.eia.gov/consumption/manufacturing/data/2018/#r1** (accessed 14 August 2024).
25 The European Chemical Industry Council (Cefic), available at:**https://cefic.org/a-pillar-of-the-european-economy/facts-and-figures-of-the-european-chemical-industry/energy-consumption/** (accessed 14 August 2024).
26 A. G. Parvatker and M. J. Eckelman, *ACS Sustainable Chem. Eng.*, 2019, **7**, 350–367.
27 J. Kleinekorte, J. Kleppich, L. Fleitmann, V. Beckert, L. Blodau and A. Bardow, *ACS Sustainable Chem. Eng.*, 2023, **11**, 9303–9319.
28 A. L. Lavoisier, *Traité élémentaire de chimie*, Paris, Cuchet, 1793.
29 *Ullmann's Encyclopedia of Industrial Chemistry Online*, ed. C. Ley, Wiley-VCH Verlag GmbH & Co. KGaA, 2002.
30 R. Bretz and P. Frankhauser, *Int. J. Life Cycle Assess.*, 1996, **1**, 139–146.
31 C. Jiménez-González, S. Kim and M. R. Overcash, *Int. J. Life Cycle Assess.*, 2000, **5**, 153–159.
32 Royal Society of Chemistry, 2024. THE MERCK INDEX online, available at:**https://merckindex.rsc.org/** (accessed 14 August 2024).
33 Elsevier B.V., 2024. Reaxys®, available at:**https://www.reaxys.com/#/search/quick** (accessed 14 August 2024).
34 European Commission, BAT reference documents, **https://eippcb.jrc.ec.europa.eu/reference** (accessed 14 August 2024).
35 E. Huber, V. Bach, P. Holzapfel, D. Blizniukova and M. Finkbeiner, *Sustainability*, 2022, **14**, 3161.
36 D. Cespi, F. Passarini, E. Neri, I. Vassura, L. Ciacci and F. Cavani, *J. Cleaner Prod.*, 2014, **69**, 17–25.
37 A. A. Adeoye, F. Passarini, J. De Maron, T. Tabanelli, F. Cavani and D. Cespi, *ACS Sustainable Chem. Eng.*, 2023, **11**, 17355–17370.
38 A. Piazzi, T. Tabanelli, A. Gagliardi, F. Cavani, C. Cesari, D. Cespi, F. Passarini, A. Conversano, F. Viganò, D. Di Bona and R. Mazzoni, *Sustainable Chem. Pharm.*, 2023, **35**, 101222–101245.
39 S. Kim and M. Overcash, *J. Chem. Technol. Biotechnol.*, 2003, **78**, 995–1005.
40 H.-J. Althaus, M. Chudacoff, R. Hischier, N. Jungbluth, M. Osses and A. Primas, *Life Cycle Inventories of Chemicals. ecoinvent report No. 8, v2.0*, EMPA Dübendorf, Swiss Centre for Life Cycle Inventories, Dübendorf, CH, 2007.
41 J. Andraos, *ACS Sustainable Chem. Eng.*, 2016, **4**, 312–323.
42 R. Cucciniello, D. Cespi, M. Riccardi, E. Neri, F. Passarini and F. M. Pulselli, *Green Chem.*, 2023, **25**, 5922–5935.
43 J. Gimenez, B. Bayarri, O. Gonzalez, S. Malato, J. Peral and S. Esplugas, *ACS Sustainable Chem. Eng.*, 2015, **3**(12), 3188–3196.
44 E. Rossi, D. Cespi, I. Maggiore, L. Setti and F. Passarini, *Environ. Sci. Pollut. Res.*, 2024, DOI: **10.1007/s11356-024-34068-1**.

45 C. Jiménez-González, M. R. Overcash and A. Curzons, *J. Chem. Technol. Biotechnol.*, 2001, **76**, 707–716.
46 D. Cespi, E. S. Beach, T. E. Swarr, F. Passarini, I. Vassura, P. J. Dunn and P. T. Anastas, *Green Chem.*, 2015, **17**, 3390–3400.
47 G. Wernet, S. Papadokonstantakis, S. Hellweg and K. Hungerbühler, *Green Chem.*, 2009, **11**, 1826–1831.
48 G. Wernet, S. Hellweg and K. Hungerbühler, *Int. J. Life Cycle Assess.*, 2012, **17**, 720–728.
49 G. Wernet, S. Hellweg, U. Fischer, S. Papadokonstantakis and K. Hungerbühler, *Environ. Sci. Technol.*, 2008, **42**, 6717–6722.
50 R. Song, A. A. Keller and S. Suh, *Environ. Sci. Technol.*, 2017, **51**, 10777–10785.
51 Y. Sun, X. Wang, N. Ren, Y. Liu and S. You, *Environ. Sci. Technol.*, 2023, **57**, 3434–3444.
52 R. Calvo-Serrano, M. González-Miquel and G. Guillén-Gosálbez, *ACS Sustainable Chem. Eng.*, 2019, **7**, 3575–3583.
53 R. Calvo-Serrano, M. González-Miquel, S. Papadokonstantakis and G. Guillén-Gosálbez, *Comput. Chem. Eng.*, 2018, **108**, 179–193.
54 P. Baxevanidis, S. Papadokonstantakis, A. Kokossis and E. Marcoulaki, *AIChE J.*, 2022, **68**, 17544.
55 J. Kleinekorte, L. Fleitmann, M. Bachmann, A. Kätelhön, A. Barbosa-Póvoa, N. von der Assen and A. Bardow, *Annu. Rev. Chem. Biomol. Eng.*, 2020, **11**, 203–233.
56 X. Zhu, C.-H. Ho and X. Wang, *ACS Sustainable Chem. Eng.*, 2020, **8**, 11141–11151.
57 H. Zarafshani, P. Watjanatepin, K. Navare, G. Sauve and K. Van Acker, *Int. J. Life Cycle Assess.*, 2024, **29**, 632–651.
58 D. Zhang, Z. Wang, C. Oberschelp, E. Bradford and S. Hellweg, *ACS Sustainable Chem. Eng.*, 2024, **12**, 2700–2708.
59 P. G. Jessop and A. R. MacDonald, *Green Chem.*, 2023, **25**, 9457–9463.
60 S&P Global Inc., Process Economics Program, available at: **https://www.spglobal.com/commodityinsights/en/ci/products/chemical-technology-pep-index.html** (accessed 14 August 2024).
61 C. Jiménez-González and M. Overcash, *Environ. Sci. Technol.*, 2000, **34**(22), 4789–4796.
62 Y. Yao and E. Masanet, *J. Cleaner Prod.*, 2018, **172**, 768–777.
63 C. F. Blanco, S. Cucurachi, F. Dimroth, J. B. Guinée, W. J. G. M. Peijnenburg and M. G. Vijver, *Energy Environ. Sci.*, 2020, **13**, 4280–4291.
64 *PRé Consultants, SimaPro v.9.6.0.1*, Amersfoort, Netherland, 2024.
65 *Ecoinvent Centre (formerly Swiss Centre for Life Cycle Inventories), Ecoinvent v.3.10 Database*, 2024.
66 Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021. The Physical Science Basis, Sixth Assessment Report*, Cambridge University Press, UK, 2021, available at: **https://www.ipcc.ch/report/ar6/wg1/** (accessed 14 August 2024).
67 R. Frischknecht, N. Jungbluth, H.-J. Althaus, C. Bauer, G. Doka, R. Dones, R. Hischier, S. Hellweg, S. Humbert, T. Köllner, Y. Loerincik, M. Margni and T. Nemecek, *Implementation of Life Cycle Impact Assessment Methods, ecoinvent report No. 3, v2.0*, Swiss Centre for Life Cycle Inventories, Dübendorf, 2007.
68 International Organization for Standardization (ISO) (2018) 14067:2018, Greenhouse gases—Carbon footprint of products—Requirements and guidelines for quantification. Geneva, Switzerland: ISO.
69 eurostat 2023, Final energy consumption in industry – detailed statistics, available at: **https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Final_energy_consumption_in_industry_-_detailed_statistics** (accessed 14 August 2024).
70 G. Pavarelli, J. Velasquez Ochoa, A. Caldarelli, F. Puzzo, F. Cavani and J.-L. Dubois, *ChemSusChem*, 2015, **8**, 2250–2259.
71 L. Cullen, F. Meng, R. Lupton and J. M. Cullen, *Nat. Chem. Eng.*, 2024, **1**, 311–322.
72 R. Heijungs, H. Udo de Haes, P. White and J. Golden, *UNEP LCA Training Kit Module k – Uncertainty in LCA*, 2008.
73 JCGM 100:2008, GUM 1995 with minor corrections Evaluation of measurement data – Guide to the expression of uncertainty in measurement.
74 UNI 11698:2017, Gestione Ambientale di Prodotto – Stima, dichiarazione e utilizzo dell'incertezza dei risultati di una Valutazione di Ciclo di Vita – Requisiti e linee guida.
75 T. Langhorst, B. Winter, D. Roskosch and A. Bardow, *ACS Sustainable Chem. Eng.*, 2023, **11**, 6600–6609.