Titre: Title:	Amélioration des bases de données d'inventaires de cycle de vie par hybridation avec l'Input Output
Auteur: Author:	Maxime Agez
Date:	2020
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Agez, M. (2020). Amélioration des bases de données d'inventaires de cycle de vie par hybridation avec l'Input Output [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5462/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/5462/
Directeurs de recherche: Advisors:	Réjean Samson, Manuele Margni, & Guillaume Majeau-Bettez
Programme: Program:	Génie chimique

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Amélioration des bases de données d'inventaires de cycle de vie par hybridation avec l'Input Output

MAXIME AGEZ

Département de génie chimique

Thèse présentée en vue de l'obtention du diplôme de *Philosophiae Doctor*

Génie chimique

Septembre 2020

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Amélioration des bases de données d'inventaires de cycle de vie par hybridation avec l'Input Output

présentée par **Maxime AGEZ**

en vue de l'obtention du diplôme de *Philosophiae Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Robert LEGROS, président

Réjean SAMSON, membre et directeur de recherche

Manuele MARGNI, membre et codirecteur de recherche

Guillaume MAJEAU-BETTEZ, membre et codirecteur de recherche

Marcelin JOANIS, membre

Gregory NORRIS, membre externe

RÉSUMÉ

Les gouvernements et entreprises sont de plus en plus pressés par les consommateurs pour prendre des décisions durables pour la planète. Afin de pouvoir prendre des décisions informées, la quantification d'émissions générées par une activité ou un produit sur la totalité du cycle de vie est nécessaire. Deux techniques principales peuvent être utilisées : l'Analyse du Cycle de Vie (ACV) et l'analyse Input-Output (IO). Ces deux techniques remplissent deux fonctions différentes. L'ACV permet de répondre à des questions précises en comparant les impacts sur l'environnement d'un produit à un autre, étant ainsi utile dans le domaine de l'écoconception. L'IO en revanche, permet de répondre à des questions de plus grosse envergure, s'intéressant davantage au système complet qu'est l'économie entière et aux interactions entre secteurs économiques. Chacune de ces méthodes toutefois, ne remplit pas son rôle à la perfection. L'ACV, qui se concentre sur la chaine de valeur d'un produit, ne parvient pas à capturer la totalité des intrants rentrant en jeu dans la chaine de valeur, menant à des sous-estimations récurrentes. L'IO, quant à elle, s'impose une agrégation de ses produits en secteurs industriels, ce qui la rend inadaptée pour étudier des produits spécifiques, diminuant la pertinence de son application.

Les forces de ces méthodes ont été combinées, donnant l'Analyse du Cycle de Vie Hybride (ACVH), qui se veut être une méthode d'évaluation du cycle de vie pouvant étudier des produits spécifiques tout en incluant l'économie entière dans la chaine de valeur de ces produits. Plus spécifiquement, l'ACVH a, en premier lieu, été développée pour compléter les intrants manquants de l'ACV avec les données de l'IO, réduisant ainsi le problème inhérent de sous-estimations de l'ACV. L'ACVH a ainsi été appliquée, de façon sporadique, pendant des décennies. Plusieurs études d'ACVH estiment ainsi de 20 à 50% l'augmentation de l'impact sur le réchauffement climatique d'un procédé ACV moyen après avoir été complété par l'IO. Cependant, l'application actuelle de l'ACVH est toujours susceptible à des sous-estimations récurrentes à cause de sa dépendance aux bases de données ACV qui ne sont pas elles-mêmes complétées par l'IO. De plus, certains aspects méthodologiques de l'ACVH ne sont pas encore entièrement pris en compte : l'effet de la correction du double comptage en ACVH n'a pas été étudié/documenté ; le manque d'opérationnalisation de l'hybridation pour des bases de données entières de façon viable sans sacrifier la qualité ; le fait que les capitaux ne soient typiquement pas endogénéisés en IO ce qui entraine également des sous-estimations ; l'absence de données de prix empêche complètement

l'hybridation ; la différence de couverture de types d'émissions entre l'ACV et l'IO montre que l'IO peut potentiellement sous-estimer les impacts environnementaux ce qui se répercuterait dans les résultats de l'ACVH.

Pour répondre à ces différentes limites identifiées, ce projet de recherche vise à développer un cadre méthodologique permettant la génération de bases de données hybrides. Cet objectif principal peut être décliné en quatre objectifs spécifiques : (1) développer une méthode de correction du double comptage en ACVH semi-automatique, i.e., une méthode harmonisant les données des deux systèmes combinés (ACV et IO) afin que chaque donnée manquante soit ajoutée, mais sans toutefois compter le même flux deux fois, (2) développer une base de données hybride pour la réalisation d'études comparatives ACVH, (3) développer une méthode d'hybridation ne dépendant pas du prix du produit de référence, et (4) intégrer l'endogénéisation des capitaux et la génération de flux élémentaires additionnels dans le cadre d'hybridation développé.

Ces objectifs spécifiques ont mené à l'écriture de trois articles scientifiques. Le premier article est une revue critique de la littérature identifiant toutes les méthodes de correction du double comptage existantes en ACVH, en les appliquant à un même exemple et comparant leurs résultats pour comprendre l'impact du choix de la méthode sur les résultats. Dans ce premier article, deux nouvelles méthodes de correction pour le double comptage sont également présentées dont une suivant des règles génériques et transparentes ce qui la rend automatisable. Le deuxième article présente le cadre d'hybridation semi-automatique développé. Ce cadre est appliqué aux bases de données ecoinvent3.5 et EXIOBASE3 et réemploie la méthode de correction du double comptage automatisée introduite dans le précédent article. L'augmentation des scores d'impact due à l'hybridation est étudiée sur quatre catégories d'impact et varie de 10 à 30% en médiane, ce qui constitue à la fois une estimation représentative des sous-estimations de l'ACV et une première étude de ces sous-estimations sur des indicateurs environnementaux autres que le réchauffement climatique. Le dernier article de ce projet répond aux limites identifiées dans l'article précédent en reprenant la méthodologie du précédent article et en y introduisant plusieurs nouveaux concepts pour perfectionner la base de données hybride générée : (1) une version d'EXIOBASE avec capitaux endogénéisés est intégrée dans le cadre d'hybridation et son impact sur les résultats étudiés, (2) des extensions environnementales additionnelles sont ajoutées à EXIOBASE et leur effet est également étudié, finalement (3) une méthode d'hybridation ne dépendant pas du prix du produit de référence est proposée. Cette nouvelle hybridation permet de contourner les problèmes posés par le manque de données de prix et peut potentiellement fournir de meilleurs résultats que l'hybridation classique. Néanmoins, cette nouvelle hybridation ne constitue qu'une proposition de méthodologie qui doit être peaufinée. Les scores d'impact issus de la base de données hybride de ce troisième article se sont concentrés sur treize catégories d'impact en utilisant la méthode d'impact IMPACT World+. Les sous-estimations dues aux données manquantes de l'ACV (ou troncatures) se sont donc vues quantifiées sur neuf nouvelles catégories et ont été complétées pour les quatre catégories déjà étudiées dans le deuxième article, menant à une estimation de la troncature variant de 1 à 200% dépendamment de la catégorie d'impact.

Les principales limites de ce projet sont liées à : l'adaptabilité du projet, qui, même si le cadre lui-même est indépendant des bases de données utilisées, a besoin de concordances que l'utilisateur doit générer entre les bases de données à utiliser ; l'impossibilité du cadre d'hybridation développé d'opérer avec une version non allouée de la base de données ACV à hybrider ; le manque d'étude d'incertitudes dans ce projet.

Ultimement, ce projet a contribué à (1) l'opérationnalisation de l'ACVH, en donnant accès à un cadre d'hybridation réutilisable pour les scientifiques, (2) la pertinence de l'ACVH, en donnant une estimation des sous-estimations inhérentes de l'ACV, (3) la performance de l'ACVH, en réglant certains problèmes méthodologiques de l'ACVH (comment hybrider avec une donnée de prix manquante, comment intégrer l'endogénéisation des capitaux en ACVH), ainsi qu'en créant une base de données hybride pour les études comparatives d'ACVH, (4) la transparence de l'ACVH, grâce à une meilleure compréhension du double comptage en ACVH, au moyen d'une étude critique des méthodes existantes et le développement de deux nouvelles méthodes.

ABSTRACT

The pressure from consumers to take durable actions for the planet on governments and companies is steadily increasing. To be able to take these actions, emissions linked throughout the lifecycle of a product or service must be quantified. Two techniques are available: Life Cycle Assessment (LCA) and input-Output (IO). Each of these methods have different goals. LCA strives to precisely compare value chains of products and is therefore greatly used in eco-design. On the other hand, IO strives to study the economy in its entirety and how sectors of the economy interact with each other. Both methods, however, do not perform their duties perfectly. LCA while striving to precisely depict the value chain of a product is unable to capture the entirety of inputs that are part of the value chain. This leads to recurring underestimations. IO struggles with the aggregation of its products into sectors which makes it highly incapable of studying specific products, thus limiting its application.

Strengths and weaknesses of these methods were combined to give Hybrid Life Cycle Assessment (HLCA). HLCA is another method quantifying lifecycle emissions that strives to fill the role of LCA and IO simultaneously. HLCA was initially designed to complete missing inputs of LCA processes with IO data. It reduces the amount of inherent underestimations in LCA. HLCA was thus applied sporadically for decades. Many HLCA studies estimate a 20 to 50% increase in global warming potential after an average LCA process is hybridized, i.e. completed with IO data. However, current application of HLCA is still prone to recurring underestimations due to its dependency to LCA databases which are not hybridized. Moreover, some methodological aspects of HLCA have yet to be fully integrated: the effect of the correction of double counting incidents in HCLA is not studied/documented; capital goods not being endogenized in IO leads to underestimations; the unavailability of price data prevents hybridization; the difference in coverage of types of emissions could lead to underestimations in impact assessment through IO which would then pass on HLCA results.

To tackle the identified issues, this project strives to develop a methodological framework able to generate hybrid databases. This main objective can be depicted into four specific objectives: (1) develop an automated method to correct double counting incidents in HLCA, (2) develop a hybrid database to be used in comparative studies, (3) develop a new hybridization method enabling the hybridization of process without price data, and (4) implement the endogenization of capital goods

and the generation of additional environmental extensions in the previously developed hybridization framework.

These specific objectives have led to three scientific articles. The first article consists in a literature review listing all existing methods to correct double counting incidents in HLCA, applying them to a same case study and analyzing the difference in the results. Two novel methods to correct double counting incidents were also introduced, one of which follows generic and transparent rules and could thus be automated. The second article presents the semi-automatic hybridization framework developed. It is then applied to ecoinvent3.5 and EXIOBASE3 databases, and uses the automated double counting correction method introduced in the first article. The increase in impacts due to hybridization was studied on four impact categories and varied from 10 to 30%. This analysis constitutes a representative estimation of underestimations in LCA as well as a first estimation of these underestimations on impact categories that are not global warming. Finally, limitations of the present hybrid database were listed. The last article builds on the results from the second by answering the identified limitations. Three new concepts were introduced: (1) a version of EXIOBASE with capital goods endogenized was integrated into the hybridization framework and its impacts on results studied; (2) additional environmental extensions were added to EXIOBASE and their effect on results were also studied; finally (3) a non-functional-flow-based hybridization method was developed, enabling the hybridization when price data is problematic. This novel hybridization method, however, only constitutes a first draft and requires additional work to identify its best use and further understands its functioning. Results of the third article were focused on thirteen impact categories, using the impact assessment method IMPACT World+. Typical underestimations in LCA could thus be quantified for nine new impact categories and were further completed for the four categories that were already included in the second article. After hybridization, impact scores of LCA were increased by 1 to 200% depending on the impact category.

The limitations of this project include: the limited adaptability of the project when changing databases used (e.g., using GTAP instead of EXIOBASE); the inability of the hybridization framework to operate with non-allocated versions of LCA databases; the lack of uncertainty analysis in this project.

Ultimately, this project contributed to (1) the operationalization of HLCA, through the hybridization framework which is reusable by scientists, (2) the relevancy of HLCA, through the representative estimation of underestimations in LCA, (3) the performance of HLCA, though the tackling of methodological issues in HLCA (notably the hybridization when price data is unavailable and the integration of endogenization of capital goods in HLCA), as well as delivering a hybrid database for comparative studies, (4) a better comprehension of double counting incidents and their correction in HLCA, through a literature review of existing methods and the development of two novel methods.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

ACV/LCA	Analyse du Cycle de Vie
ACVH/HLCA	Analyse du Cycle de Vie Hybride
EEIO	Environmentally Extended Input-Output
IO	Input-Output
IW+	IMPACT World+

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

Notre société a un impact sur la planète. Toutes nos activités que ce soit utiliser sa voiture, partir en vacances, surfer sur le Net ou tout simplement respirer entrainent des conséquences plus ou moins importantes sur notre planète, son environnement et ses ressources (IPCC, 2014). Par exemple, chaque année on mesure le jour du dépassement qui représente la date à partir de laquelle le cout de nos activités sur la planète dépasse ce qu'elle est capable de régénérer en une année. Autrement dit, c'est la date à partir de laquelle notre société vit à crédit. En 1986, ce jour du dépassement était le 30 octobre alors qu'en 2019 c'était le 29 juillet (Earth Overshoot Day, 2019), indiquant une nette augmentation de la consommation des ressources. Dans la même veine, certains indicateurs écologiques comme le réchauffement de la planète ou la perte de biodiversité, menacent également de créer des changements drastiques pour la société (Rockström et al., 2009). Pour ne citer que quelques exemples *possibles* qui nous attendent : catastrophes climatiques, compétition entre consommateurs pour accéder aux ressources, déplacements de population massifs (et les problèmes qui y sont liés), l'extinction de plusieurs espèces chaque jour. Chaque acteur, i.e., gouvernements, entreprises, consommateurs, doit donc chercher à réduire son propre impact en changeant leurs actions actuelles. Mais comment savoir quelles actions sont à changer car mauvaises pour l'environnement ou au contraire à conserver car déjà bonnes pour l'environnement ?

En tant que gouvernement, dois-je prôner le développement de la mobilité électrique dans mon pays ? Quels impacts cela aura-t-il sur l'empreinte écologique liée à mon pays et mes concitoyens ? Dans quelle mesure mon choix impactera-t-il la planète dans son ensemble et tous ses habitants ?

En tant qu'industriel, quel fournisseur me donne accès aux produits les moins polluants ? Comment puis-je diminuer mon empreinte écologique afin de répondre aux attentes de mes investisseurs / afin de répondre aux législations en vigueur ou à venir ? Comment devenir un leadeur de mon secteur en matière d'environnement ?

En tant que consommateur, comment puis-je faire une différence à mon échelle ? Dois-je acheter des tomates locales ou importées ? Est-ce mieux de se convertir au végétarisme ou d'abandonner un voyage en vacances par an ?

Pour répondre à toutes ces questions, il faut être en mesure de quantifier les émissions associées à la production, distribution, utilisation et fin de vie (i.e., du cycle de vie) de tout produit ou service, ainsi que l'impact de ces émissions sur l'environnement. Il faut également veiller à regarder l'ensemble du cycle de vie afin de ne pas déplacer un problème dans une autre phase de cycle de vie au lieu de régler ledit problème (UNEP/SETAC Life Cycle Initiative, 2012). Par exemple, ne regarder que l'étape d'utilisation d'une voiture électrique et conclure qu'elle est automatiquement meilleure qu'une voiture à essence parce qu'elle n'émet que très peu d'émissions lors de son utilisation est incorrect. Une fois qu'on a également inclus la production de l'électricité utilisée pour alimenter la batterie de la voiture électrique (qui peut être générée à partir de charbon dépendamment des pays) ainsi que tout le reste du cycle de vie, on pourra alors conclure si l'utilisation d'une voiture électrique dans le contexte particulier de l'utilisateur est meilleure ou non qu'utiliser une voiture à essence.

Pour ce faire, deux méthodes sont actuellement principalement utilisées : L'Analyse de Cycle de Vie (ACV) et l'analyse Input-Output (IO). Ces méthodes reposent sur l'utilisation de millions de données contenues dans des bases de données, générées par de nombreux experts dans le monde. Ce sont ces bases de données qui assurent la couverture de tout le cycle de vie de chaque produit, car elles contiennent des informations sur la plupart des produits/services existants dans notre société. Plus spécifiquement, combien de ressources, électricité, infrastructure, etc. sont nécessaires pour produire tel ou tel produit, et combien d'émissions sont dégagées lors de la production. On parle de « recettes technologiques » (ou de fonctions de production) (Leontief, 1970), tout comme pour faire un gâteau il faut une certaine proportion de tel et tel ingrédient.

Ces méthodes sont d'ores et déjà utilisées dans la détermination d'impacts sur l'environnement par de nombreux gouvernements/entreprises (Sonnemann, Gemechu Eskinder, Sala, & Schau, 2017). Cependant, de façon isolée, elles ne sont pas capables de répondre parfaitement aux questions posées par la société. L'ACV n'est pas en mesure de donner des renseignements sur chaque produit/service existant dans notre société (e.g., pas de recette pour « faire une transaction bancaire »), mais peut renseigner sur des produits spécifiques (e.g., tomates vs bananes). Par conséquent, l'ACV n'est pas capable de donner une estimation complète des émissions car elle manque de données sur certaines activités de notre société, ce qui *peut* mener à de mauvaises conclusions (Lenzen, 2000). L'IO quant à elle, peut donner des renseignements sur tous les produits/services existants, mais ne donne accès qu'à des informations agrégées (e.g., fruits). Elle peut donc donner des estimations complètes mais uniquement sur des secteurs économiques et ne peut donc pas répondre à des questions sur des produits précis (Suh et al., 2004).

Heureusement, leur force et faiblesse sont complémentaires, elles ont donc pu être combinées en ce qu'on appelle l'Analyse du Cycle de Vie Hybride (ACVH) (Bullard & Penner, 1978). L'ACVH a été créée dans le but de parfaire les réponses apportées par l'ACV et l'IO, c'est-à-dire une estimation des émissions complète et disponible pour des produits spécifiques (i.e., recette disponible pour une tomate, une banane ou un avocat et non seulement pour un fruit). Toutefois, elle-même connait des limites dans son application et dans la qualité des données des bases de données utilisées, ce qui l'empêche également de fournir la meilleure quantification possible d'émissions liées à une activité ou un service.

Le travail présenté dans cette thèse cherche à perfectionner la quantification d'émissions des bases de données utilisées en ACV, IO et ACVH afin de fournir les meilleures informations possibles à chaque preneur de décision. Pour ce faire l'objectif général de la thèse est de combiner bases de données ACV et IO.

CHAPITRE 2 REVUE DE LITTERATURE

2.1 Analyse environnementale du Cycle de Vie

L'Analyse environnementale du Cycle de Vie (ACV par la suite) est un outil d'aide à la décision destiné à déterminer les impacts sur l'environnement résultant du cycle de vie d'un produit ou service. Ce cycle de vie peut être divisé en cinq étapes : l'acquisition des ressources, la production, la distribution, l'utilisation et la fin de vie (voir figure 4-1).

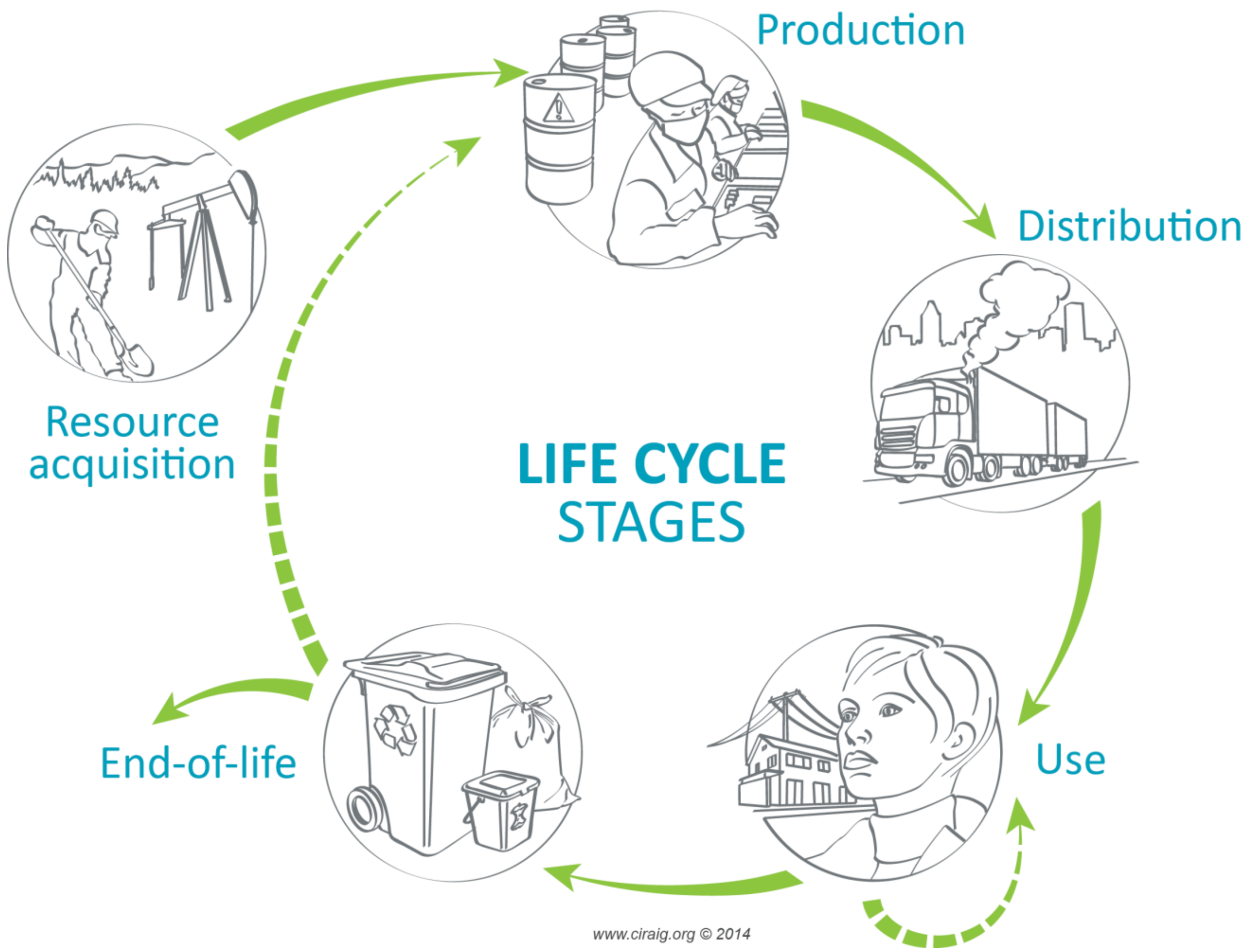


Figure 2-1 Les cinq étapes du cycle de vie (tiré de http://ciraig.org)

La pensée cycle de vie cherche à inclure en tout temps ces cinq étapes ainsi que toutes les chaines de valeur entrant dans la production d'un bien ou service, afin d'éviter un déplacement d'impact. En d'autres termes, l'ACV cherche à éviter de réduire les émissions sur une partie du cycle de vie tout en les augmentant sur une autre partie (UNEP/SETAC Life Cycle Initiative, 2012). Tous matériaux et émissions rentrant dans le cycle de vie d'un bien ou service sont donc inclus dans ce qu'on appelle un inventaire de cycle de vie. Cette pratique remonte à 1963 lorsque Harold Smith réalisa un inventaire cumulatif de la demande énergétique liée à un procédé chimique (Smith, 1963). Depuis, la méthodologie ACV a évolué, est devenue reconnue et son application est soumise à des normes (famille des normes ISO14040… plus spécifiquement) (International Organization for Standardization, 2006). Cette méthodologie suit quatre étapes :

1. <u>Objectifs et champ de l'étude</u> : cette étape sert à définir les questions auxquelles l'étude ACV cherche à répondre. On y définit donc les objectifs de l'étude et plusieurs paramètres importants (unité fonctionnelle, frontières du système, méthodes d'allocation, catégories d'impact environnemental, etc.). Le paramètre le plus important étant l'unité fonctionnelle permettant de comparer les performances de deux systèmes répondant à une même fonction, e.g., « transporter 1 passager sur 1000km ». Grâce à cette unité fonctionnelle, on peut comparer les performances d'une voiture et d'un avion sur la base d'une fonction équivalente.
2. <u>Inventaire du cycle de vie</u> : cette étape cherche à quantifier tous les flux économiques (i.e., flux provenant de l'économie) et tous les flux élémentaires (i.e., flux provenant de/aboutissant dans l'environnement) entrants et sortants du système défini par les frontières du système. Ces frontières représentent la délimitation entre les données incluses ou non dans l'étude. On obtient donc à la fin un inventaire de tous les matériaux et émissions nécessaires pour répondre à l'unité fonctionnelle posée (illustration d'un inventaire de cycle de vie en figure 2-2).
3. <u>Evaluation des impacts du cycle de vie</u> : dans cette étape on vient traduire les résultats de l'inventaire du cycle de vie en impact sur l'environnement. Ces impacts sont séparés en plusieurs catégories d'impact (e.g., réchauffement climatique, acidification, eutrophisation, etc.). Ils sont déterminés en utilisant des facteurs de caractérisation et les impacts peuvent être étudiés en tant que problèmes (aussi appelé midpoint) ou dommages (aussi appelé

endpoint). L'utilisation d'impacts dommages facilite la compréhension des résultats, étant donné que l'on passe d'une douzaine d'indicateurs problèmes à trois indicateurs dommage (santé humaine, qualité des écosystèmes et utilisation des ressources). En revanche, l'incertitude autour de ces impacts dommages est plus grande.

4. <u>Interprétation des résultats</u> : les résultats obtenus sont analysés pour en tirer des conclusions. Il est très rare qu'un système étudié soit meilleur qu'un autre sur la totalité des impacts pris en compte. L'étape d'interprétation des résultats cherche donc à identifier pour le décideur les différents compromis existants.

Les résultats de l'ACV sont déterminés grâce à l'équation suivante (adaptée de (Heijungs & Suh, 2002)) :

$$\boldsymbol{q}^{lca} = \boldsymbol{C}^{lca} \cdot \boldsymbol{B}^{lca} \cdot (\boldsymbol{I} - \boldsymbol{A}^{lca})^{-1} \cdot \boldsymbol{y}^{lca} \qquad (1)$$

où $\boldsymbol{q}^{lca}$ sont les impacts sur l'environnement, $\boldsymbol{C}^{\boldsymbol{lca}}$ est une matrice contenant les facteurs de caractérisation utilisés pour traduire les différentes émissions en impact sur l'environnement, $\boldsymbol{B}^{lca}$ est une matrice contenant les flux élémentaires de chaque procédé inclus dans le cycle de vie, $\boldsymbol{I}$ est la matrice identité (i.e., avec des 1 sur toute la diagonale), $\boldsymbol{A}^{lca}$ est la matrice contenant tous les flux économiques (e.g., matériaux requis)[1], $\boldsymbol{y}^{lca}$ représente l'unité fonctionnelle, dans laquelle on demande une certaine quantité des processus élémentaires étudiés. Une représentation visuelle de ces matrices est donnée en figure 2-2. On y voit que la production d'une voiture X (fictive) requiert 0.5kWh d'électricité, 2kg d'acier, etc. et émet 0.3kg de CO_2 directement.

[1] Dans l'équation 1, la convention IO est utilisée. La matrice A inclut donc uniquement les usages normalisés des biens et non la production du bien lui-même qui est retrouvée dans la matrice identité.

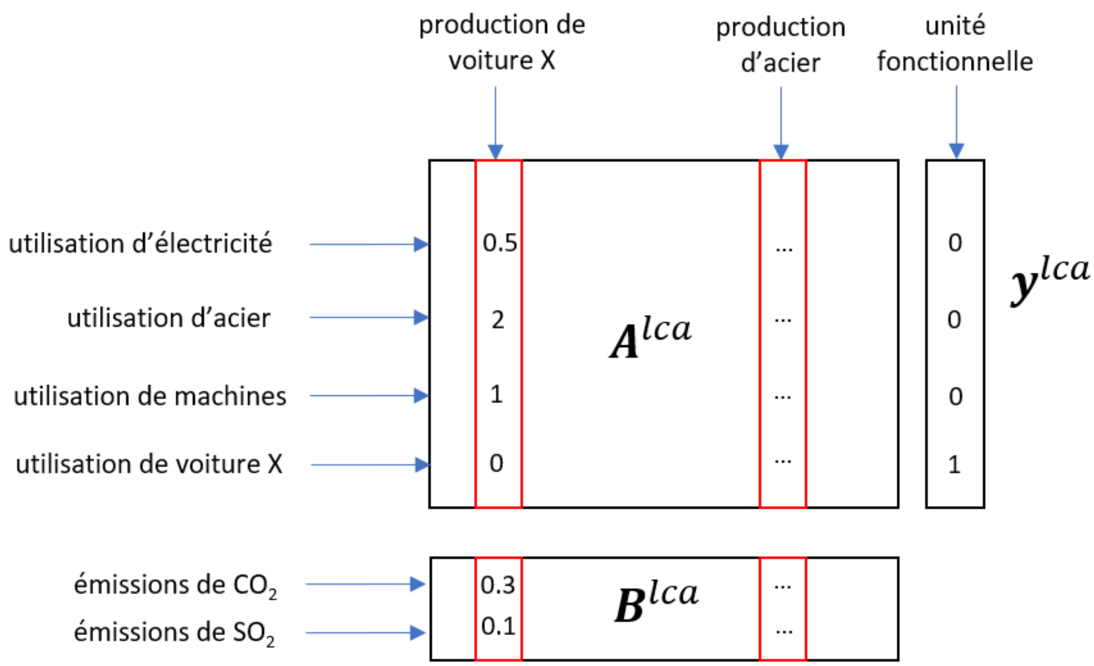


Figure 2-2 Représentation visuelle des matrices utilisées par l'ACV

Un inventaire du cycle de vie typique inclut des millions de données car non seulement l'électricité utilisée dans la production du produit étudié doit être incluse mais également le charbon utilisé pour produire cette électricité, puis les camions qu'il a fallu pour extraire ce charbon, et ainsi de suite. On se trouve rapidement avec un système de millions de données à collecter (figure 2-3), ce qui n'est pas concevable d'un point de vue ressources (temps et argent). L'ACV repose donc sur des bases de données qui contiennent la modélisation de procédés génériques, comme la production d'électricité (Wernet et al., 2016). Le praticien ACV peut donc réutiliser cette modélisation de l'électricité et a juste à déterminer combien d'électricité le procédé étudié consomme. On a donc des données dites « d'avant-plan » qui sont générées par le praticien ACV lui-même (e.g., la quantité d'électricité consommée) et des données dites « d'arrière-plan » qui proviennent de bases de données (e.g., combien de charbon, etc. requis pour fournir de l'électricité). Une base de données est donc une collection de procédés génériques qu'on appelle processus élémentaires. La figure 2-4 montre un de ces processus élémentaires, tel que présent dans le logiciel ACV openLCA. On peut y voir les flux économiques nécessaires : « glider », « engine », etc.

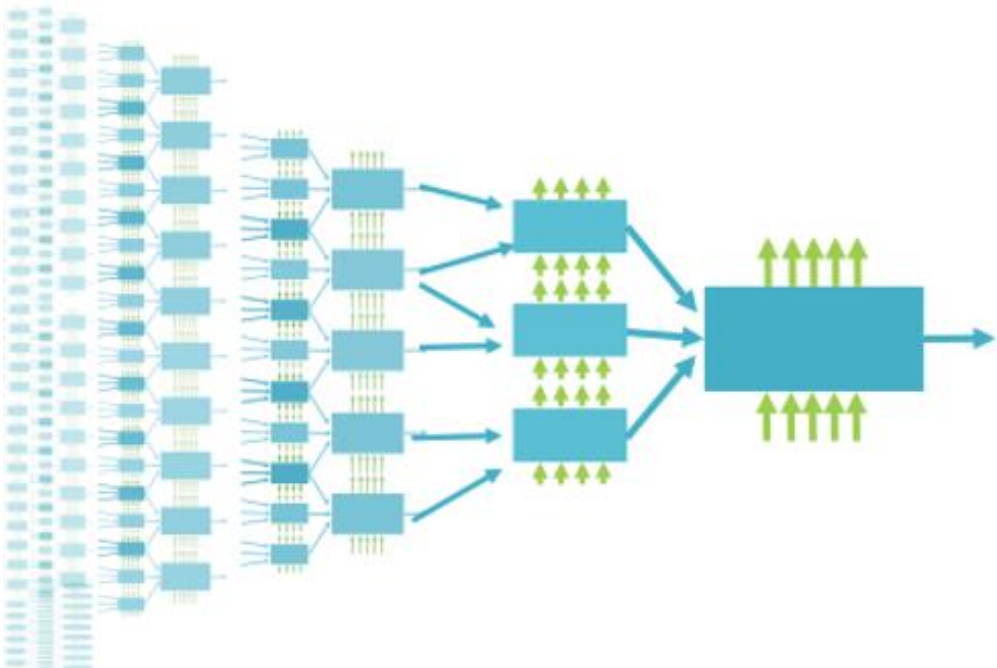
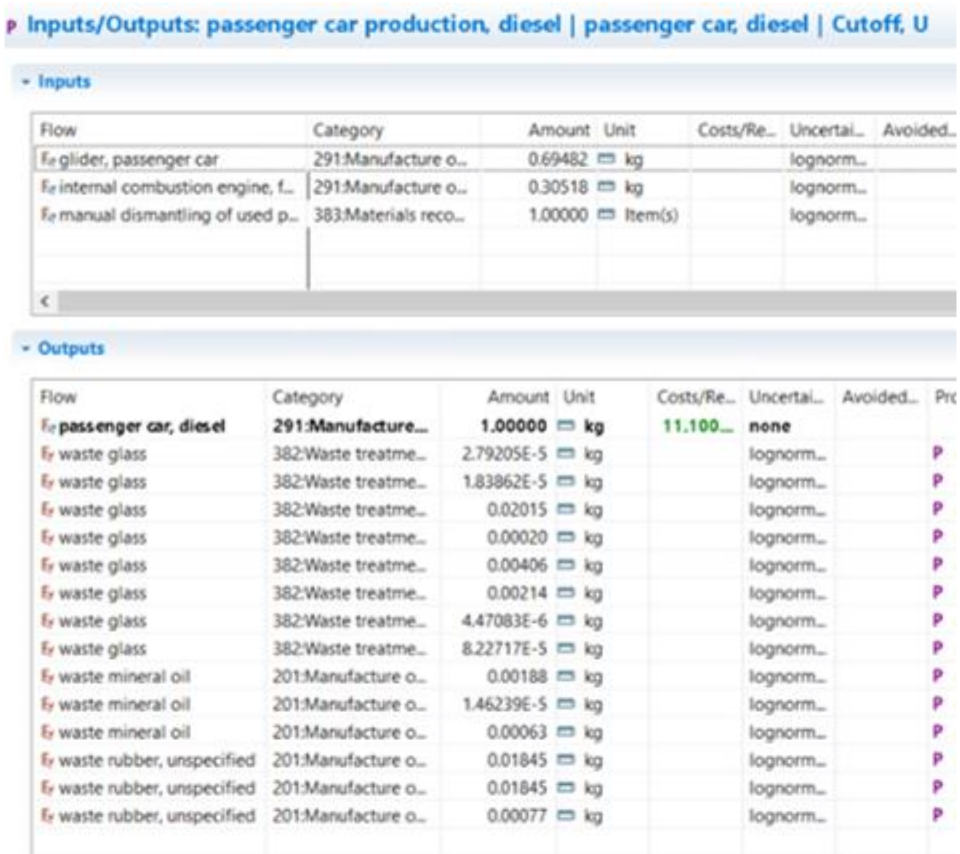


Figure 2-3 Représentation de l'inventaire du cycle de vie d'un produit/service (image fournie par Pascal Lesage)



Inputs/Outputs: passenger car production, diesel | passenger car, diesel | Cutoff, U

Inputs

Flow	Category	Amount	Unit	Costs/Re...	Uncertai...	Avoided...
glider, passenger car	291:Manufacture o...	0.69482	kg		lognorm...	
internal combustion engine, f...	291:Manufacture o...	0.30518	kg		lognorm...	
manual dismantling of used p...	383:Materials reco...	1.00000	Item(s)		lognorm...	

Outputs

Flow	Category	Amount	Unit	Costs/Re...	Uncertai...	Avoided...	Pr
passenger car, diesel	**291:Manufacture...**	**1.00000**	**kg**	11.100...	**none**		
waste glass	382:Waste treatme...	2.79205E-5	kg		lognorm...		P
waste glass	382:Waste treatme...	1.83862E-5	kg		lognorm...		P
waste glass	382:Waste treatme...	0.02015	kg		lognorm...		P
waste glass	382:Waste treatme...	0.00020	kg		lognorm...		P
waste glass	382:Waste treatme...	0.00406	kg		lognorm...		P
waste glass	382:Waste treatme...	0.00214	kg		lognorm...		P
waste glass	382:Waste treatme...	4.47083E-6	kg		lognorm...		P
waste glass	382:Waste treatme...	8.22717E-5	kg		lognorm...		P
waste mineral oil	201:Manufacture o...	0.00188	kg		lognorm...		P
waste mineral oil	201:Manufacture o...	1.46239E-5	kg		lognorm...		P
waste mineral oil	201:Manufacture o...	0.00063	kg		lognorm...		P
waste rubber, unspecified	201:Manufacture o...	0.01845	kg		lognorm...		P
waste rubber, unspecified	201:Manufacture o...	0.01845	kg		lognorm...		P
waste rubber, unspecified	201:Manufacture o...	0.00077	kg		lognorm...		P

Figure 2-4 Inventaire du processus élémentaire production d'1kg de voiture diesel de la base de données ecoinvent3.6 (capture d'écran du logiciel openLCA).

Les frontières du système définies dans l'étape « objectifs et champ de l'étude » représentent la délimitation entre les données incluses ou non dans l'étude. Dans le cas où une donnée est manquante, les praticiens ont tendance à exclure cette donnée, et donc à l'exclure des frontières du système. En d'autres termes, cette donnée, ses émissions et son impact sur l'environnement seront ignorés. Les frontières de système constituent la principale raison pour laquelle deux ACV étudiant le même système n'arrivent typiquement pas aux mêmes conclusions (Suh et al., 2004). Plus les frontières du système incluent de procédés, plus l'impact sur l'environnement augmente. Il est donc possible de volontairement réduire les frontières du système (i.e., exclure des procédés) afin de réduire les impacts environnementaux calculés, ce qui constitue une des principales critiques envers l'ACV.

Toutefois, même en cherchant à étendre au maximum les frontières du système, il y aura toujours des procédés exclus de l'analyse ultimement. La raison est simple. A l'heure de la mondialisation, vouloir faire l'ACV d'un produit, quel qu'il soit, revient à devoir inclure l'économie dans son ensemble (Gibon & Schaubroeck, 2017; Pomponi & Lenzen, 2017), ce qui n'est pas faisable avec les méthodes *actuelles* de l'ACV.

L'ACV souffre donc dans la grande majorité des cas d'un problème dit de *troncature*. Puisque les frontières du système de l'étude excluent des procédés (Suh et al., 2004), les impacts environnementaux liés à ces procédés sont systématiquement ignorés dans l'analyse. Cette omission systématique d'impacts crée logiquement un biais négatif, menant à une plus grande probabilité de sous-estimation que de surestimation de nos échanges avec l'environnement (Gibon & Schaubroeck, 2017; Pomponi & Lenzen, 2017).

Les décisions prises basées sur des résultats sous-estimés peuvent ne pas mener à l'amélioration espérée étant donné qu'il existe une part du système qui est ignorée et qui pourrait changer les conclusions de l'ACV (Lenzen & Treloar, 2003). Le domaine de l'ACV, bien que conscient de ce problème, n'a pas d'idée de la proportion du système qui est en réalité exclue des frontières du système et de son impact, et simplement considère que l'impact de ces procédés devrait être le plus négligeable possible. Il est vrai que l'impact de ces procédés regardés séparément sera, dans la majorité des cas, négligeable, mais l'impact cumulé de millions de procédés exclus ne sera quant à lui, pas forcément négligeable.

2.2 Analyse Input-Output

2.2.1 Introduction de l'analyse Input Output et de l'analyse Environmentally Extended Input Output

L'analyse Input-Output (IO par après) est un modèle économique quantitatif montrant les relations intersectorielles de l'économie d'une ou plusieurs régions, le tout relié à une demande finale (e.g., consommateurs). Dans les années 1930 et pendant la Deuxième Guerre mondiale, certains pays (dont les États-Unis) utilisaient l'IO afin de planifier au mieux l'économie, i.e., où davantage de ressources étaient nécessaires ou au contraire étaient suffisantes voire excédentaires (Dantzig, 1951; Leontief, 1936). Les relations intersectorielles sont représentées par une matrice dite technologique, regroupant les transactions financières entre secteurs. Ces transactions forment des fonctions de production décrivant la quantité d'intrants requis pour la fabrication de produits. Une appellation équivalente employée précédemment dans cette thèse est « recette technologique ». Ces fonctions de production sont typiquement en unités monétaires mais peuvent également inclure des unités physiques (Merciai & Schmidt, 2017). La demande totale sur l'économie (x) étant calculée grâce à l'équation de Leontief :

$$\boldsymbol{x} = \left(\boldsymbol{I} - \boldsymbol{A}^{io}\right)^{-1} \cdot \boldsymbol{y}^{io} \tag{2}$$

où $\boldsymbol{I}$ est la matrice identité (1 sur la diagonale), $\boldsymbol{A^{io}}$ étant la matrice technologique et $\boldsymbol{y^{io}}$ étant la demande finale, composée notamment des consommateurs et du gouvernement.

Depuis les années 70, l'IO a évoluée pour inclure des extensions environnementales dans son modèle (Leontief, 1970). On parle alors d'analyse « Environmentally Extended Input Output » (ou EEIO). Dans la suite de cette thèse, les termes IO et EEIO sont souvent interchangeables. L'EEIO constitue un autre outil d'aide à la décision respectant la pensée de cycle de vie. Contrairement à l'ACV, l'IO se concentre sur les secteurs de l'économie (et non sur des produits spécifiques), pour des raisons qui seront expliquées dans la sous-section 2.2.2. La figure 2-5 montre une fonction de production tirée de la table IO EXIOBASE. On y voit les flux économiques nécessaires à la production *moyenne* d'un euro de véhicule motorisé au Canada. Par exemple, 0.035€ de produits d'acier venant du Canada est nécessaire à la production de 1€ de véhicules motorisés au Canada. A noter que l'utilisation de secteurs d'activités entraine de la demande intrasectorielle, i.e., les pièces de la voiture sont dans le même secteur que la voiture elle-même. Les impacts du cycle de vie sur l'environnement sont déterminés grâce à l'équation suivante :

$$\boldsymbol{q}^{io} = \boldsymbol{C^{io}} \cdot \boldsymbol{B}^{io} \cdot \left(\boldsymbol{I} - \boldsymbol{A}^{io}\right)^{-1} \cdot \boldsymbol{y}^{io} \tag{3}$$

où $\boldsymbol{q}^{io}$ contient les scores d'impact sur l'environnement, $\boldsymbol{C^{io}}$ contient les facteurs de caractérisation liants flux élémentaires (e.g., CH_4) à leur impact sur l'environnement (réchauffement climatique), $\boldsymbol{B}^{io}$ est la matrice d'extensions environnementales (voir figure 2-6 pour un exemple de cette matrice), $\boldsymbol{I}$, $\boldsymbol{A}^{io}$ et $\boldsymbol{y}^{io}$ sont les mêmes que dans l'équation (2). On remarque une ressemblance dans la structure des équations utilisées en ACV et en EEIO (équations (1) et (3)).

region	sector	Motor vehicles, trailers and semi-trailers (34)
US	Motor vehicles, trailers and semi-trailers (34)	0.137305
CA	Motor vehicles, trailers and semi-trailers (34)	0.125696
	Services auxiliary to financial intermediation (67)	0.073756
	Other business services (74)	0.043855
	Basic iron and steel and of ferro-alloys and first products thereof	0.034808
	Sale, maintenance, repair of motor vehicles, motor vehicles parts, motorcycles, motor cycles parts and accessoiries	0.033509
	Foundry work services	0.023151
	Retail trade services, except of motor vehicles and motorcycles; repair services of personal and household goods (52)	0.020437
MX	Motor vehicles, trailers and semi-trailers (34)	0.018413
US	Basic iron and steel and of ferro-alloys and first products thereof	0.015692
DE	Motor vehicles, trailers and semi-trailers (34)	0.009353
CA	Fabricated metal products, except machinery and equipment (28)	0.008943
	Supporting and auxiliary transport services; travel agency services (63)	0.008138
	Rubber and plastic products (25)	0.007826
US	Rubber and plastic products (25)	0.007755
JP	Motor vehicles, trailers and semi-trailers (34)	0.007629
US	Machinery and equipment n.e.c. (29)	0.007365
	Electrical machinery and apparatus n.e.c. (31)	0.006839
CA	Financial intermediation services, except insurance and pension funding services (65)	0.006561

Figure 2-5 La liste ordonée (et tronquée par manque d'espace) des flux économiques directs requis (en €) pour la production d'un euro de véhicule motorisé au Canada. Cette liste représente une colonne de la matrice $\boldsymbol{A^{io}}$ (obtenu de la base de données EXIOBASE3).

	(CA, Motor vehicles, trailers and semi-trailers (34))	unit
CO2 - combustion - air	0.016762	kg
PFC - air	0.000199	kg
NMVOC - non combustion - Manufacture of automobiles - air	0.000193	kg
HFC - air	0.000120	kg
CO - non combustion - Agglomeration plant - sinter - air	0.000112	kg
CO - non combustion - Pig iron production, blast furnace - air	0.000086	kg
SOx - non combustion - Zinc, unwrought, not alloyed - air	0.000046	kg
TSP - non combustion - Pig iron production, blast furnace - air	0.000030	kg
CO - non combustion - Primary aluminium production - air	0.000029	kg
CO - non combustion - Steel production: basic oxygen furnace - air	0.000029	kg
SOx - non combustion - Refined copper; unwrought, not alloyed - air	0.000028	kg
SOx - non combustion - Refined lead, unwrought - air	0.000025	kg

Figure 2-6 Extensions environnementales d'EXIOBASE3 liées à la production d'un euro de véhicule motorisé au Canada (obtenu de la base de données EXIOBASE3).

2.2.2 Les hypothèses de l'IO

L'IO, dans sa version originelle, est un modèle économique non contraint, linéaire, avec des fonctions de production fixes. Qu'est-ce que cela entraine ?

2.2.2.1 La linéarité

L'IO suit une hypothèse de linéarité, c'est-à-dire que la quantité d'intrants requis augmente linéairement avec la quantité de produits fabriqués (Vogstad, 2009). Par exemple, dans la figure 2-5, on voit un intrant de « Basin iron and steel … » requis dans la fabrication d'un véhicule motorisé. Si la quantité requise de véhicules motorisés doublait, la quantité d'acier doublerait également. Il n'y a nul doute qu'une augmentation de la quantité à produire de véhicules entraine une consommation accrue d'acier. En revanche, cette consommation n'est pas forcément linéaire dans la vraie vie. Cette hypothèse de linéarité est également adoptée par l'ACV.

2.2.2.2 Des fonctions de production fixes

L'IO est basée sur des fonctions de production fixes (Koesler & Schymura, 2015). En effet, l'IO, après avoir défini un ensemble d'intrants requis pour produire une commodité (comme décrit en figure 2-5), réutilise cette même fonction de production pour tout produit fabriqué par le secteur. En d'autres termes, chaque véhicule produit par le secteur « véhicules motorisés » (de la figure 2-5) sera composé de 0.035€ d'acier par € de véhicule produit. De plus, la proportion entre les différents intrants demeure également inchangée. Ces fonctions de production fixes incluent aussi les extensions environnementales de l'EEIO.

Cette hypothèse de l'IO rend son pouvoir prospectif faible, car les fonctions de production ne se réarrangent pas suite à un changement majeur, il n'y a donc pas de concept d'économie d'échelle ni d'élasticité de prix. Introduire une élasticité de prix dans l'IO donne lieu à ce qu'on appelle les modèles d'équilibres (généraux ou partiels) (Burfisher, 2011; Wing, 2004).

2.2.2.3 Modèle non contraint

L'IO est un modèle économique non contraint (Duchin & Levine, 2011) c'est-à-dire que les réalités contextuelles, technologiques ou de ressources ne sont pas prises en compte. Par exemple, l'IO serait capable d'étudier la multiplication de la production nationale annuelle de véhicules par 100, alors que les ressources ne pourraient pas être disponibles par exemple, ou que les technologies ne seraient tout simplement pas capables de répondre à ce genre d'augmentation.

Le manque de contrainte limite encore un peu plus le pouvoir prospectif de l'IO, car ne permet pas de vérifier si la solution étudiée est faisable ou non. L'introduction de contraintes dans le modèle IO donne lieu à des modèles de programmation linéaire (ou optimisation linéaire) (Azapagic & Clift, 1998; Duchin, 2005; Duchin, Levine, & Strømman, 2016; Katelhön, Bardow, & Suh, 2016).

2.2.2.4 Des équilibres à respecter

L'IO se base sur deux équilibres. Un équilibre dit financier et un équilibre dit des marchés (R. E. Miller & Blair, 2009b). L'équilibre financier consiste à s'assurer que la valeur totale produite est égale à la somme de tous les intrants et de la valeur ajoutée. Si cet équilibre n'est pas respecté, on peut se retrouver avec des fonctions de production incohérentes, où produire 1€ de produit a couté moins (ou plus) que 1€. De façon pratique, il faut s'assurer que chaque colonne de la matrice technologique $\boldsymbol{A^{io}}$ additionnée avec leurs facteurs de production économiques respectifs (salaires, marges, etc.) donne 1.

L'équilibre des marchés constitue une nécessité comptable. D'une part, elle s'assure que tous biens consommés sont effectivement produits. D'autre part, elle s'assure que tous les biens produits ont un devenir documenté: ils sont soit consommés dans la production d'autres biens (consommation intermédiaire), consommés par les utilisateurs finaux (ménages, gouvernements, etc.), ou stockés en inventaires (ajouts nets aux stocks, ce qui en pratique est traité comme une catégorie de la demande finale).

2.2.3 La génération de données en IO

D'où proviennent toutes les données nécessaires ? La génération des données en EEIO se fait au moyen de d'enquêtes lancées par les agences statistiques nationales. Dans ces enquêtes, les entreprises de moyenne/grande taille sont tenues de déclarer leurs transactions financières annuelles, i.e., déclarer ce qu'elles ont acheté et vendu. Ces inventaires nationaux sont ensuite regroupés en des comptes nationaux, décrivant l'état de l'économie. De nombreux rééquilibrages et réconciliations de données sont nécessaires. Les données de chaque entreprise (obtenues grâce aux enquêtes) sont ensuite agrégées par secteur d'activité (European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, &

World Bank, 2009). Par exemple, une entreprise comme Renault serait rangée dans le secteur d'activité « production de véhicules ». Les entreprises sont classées par secteur afin d'éviter de donner publiquement les données (et modes de production) des entreprises.

Ces transactions sont ensuite regroupées en deux tables ressources et emplois (Supply and Use tables en anglais) (European Commission, 2008). La table des ressources décrit la vente de chaque secteur. La table des emplois décrit les acquisitions de chaque secteur. Puis, au moyen de « constructs » (e.g., industry-technology, commodity-technology) (Majeau-Bettez, Wood, & Strømman, 2014), ces deux tables sont regroupées en une matrice technologique ($\boldsymbol{A}^{io}$).

Les matrices d'extensions environnementales quant à elle, sont déterminées après coup en utilisant des facteurs d'émissions calculés à partir de statistiques nationales (Stadler et al., 2018). Ces statistiques nationales environnementales ne découlent pas des enquêtes utilisées par la comptabilité nationale qui se concentrent sur les transactions économiques.

2.2.4 Limites de l'IO

Vu le processus de génération des données, on peut distinguer plusieurs limites de l'EEIO : (1) les données sont générées à l'année. L'EEIO (et l'IO) est donc simplement un état des lieux de l'économie à une année précise. De plus, la compilation des données est longue. Les tables EEIO sortent donc avec plusieurs années de retard. (2) Les données sont uniquement sectorielles. Il est donc impossible d'évaluer l'empreinte environnementale d'un produit spécifique (e.g., tomate), puisque les entreprises produisant ce produit spécifique sont regroupées avec d'autres produits similaires, mais différents (e.g., producteurs de fraises, pommes, etc.). (3) L'EEIO est soumise à un problème de biais d'agrégation (« aggregation bias ») (Morimoto, 1970). L'agrégation de fonctions de production en secteurs, dans un certain ratio, peut entrainer des surestimations (ou sous-estimations), si ces fonctions ne sont pas utilisées dans le même ratio avec lequel elles ont été agrégées (voir annexe B pour une démonstration et explication du biais d'agrégation de l'IO). Les attributions d'impact d'études rétrospectives avec l'IO ainsi que les impacts totaux d'études prospectives/analyses de scénario sont donc biaisés (Steen-Olsen, Owen, Hertwich, & Lenzen, 2014). Cette surestimation (ou sous-estimation) est appelée biais d'agrégation. Elle constitue un problème majeur de l'IO. (4) L'agrégation des données en secteur entraine potentiellement une hétérogénéité de prix, i.e., les produits au sein d'un même secteur ont des prix variant énormément même si les produits eux-mêmes sont homogènes. Cette hétérogénéité des prix est un problème pour l'ACV hybride comme détaillé dans la section 2.4.5. (5) La fixité des fonctions de production en IO rend cette méthode statique contrairement aux modèles d'équilibre qui apportent une flexibilité, notamment avec la notion d'élasticité des prix. En revanche, cette limite n'est pas gênante pour l'application de l'IO avec l'ACV, étant donné que l'ACV elle-même se concentre sur des fonctions de production fixes. Cette même fixité des fonctions de production est observée par l'IO et l'ACV sur leur façon respective de comptabiliser les biens de capital. La rigidité de l'IO quant aux biens de capital n'est donc également pas un problème dans son application avec l'ACV. A noter que dans une logique rétrospective, comme il est souvent le cas en IO comme en ACV (dite attributionelle), le recours aux modèles d'équilibre n'est pas nécessaire.

2.3 ACV Hybride

2.3.1 A quoi sert l'ACVH ?

L'ACV Hybride (ACVH par la suite) combine ACV et IO afin d'obtenir un autre outil d'aide à la décision respectant la pensée cycle de vie. L'ACVH existe depuis plusieurs décennies (Bullard & Penner, 1978) et son but principal est de répondre aux problèmes de troncature de l'ACV et de biais d'agrégation de l'IO. L'ACVH cherche donc à être une méthode d'évaluation capable de regarder des problèmes spécifiques (comme l'ACV), mais sans problème de troncature. L'ACVH est toutefois rarement utilisée même si ces bienfaits ont été mis de l'avant. Plusieurs études comparant résultats obtenus avec l'ACVH et ACV (sur l'indicateur de réchauffement climatique) montrent que les résultats obtenus avec l'ACVH sont typiquement de 20 à 50% plus élevés (Ferrao & Nhambiu, 2009; Junnila, 2006; Lenzen, 2000; Lenzen & Dey, 2000; Norris, 2002; Rowley, Lundie, & Peters, 2009). Cette augmentation peut servir d'estimation grossière du niveau de troncature de l'ACV, même si ces études ont été réalisées par différents groupes, obtenus avec différentes méthodologies et concentrées sur des secteurs spécifiques de l'économie. Elles ne constituent donc pas une réponse définitive quant au niveau de troncature moyen de l'ACV. Néanmoins, ces études montrent que dans certains cas, les résultats obtenus avec l'ACV sont grandement sous-estimés. De manière générale, Pomponi et Lenzen (2017) montrent que l'erreur de troncature de l'ACV est typiquement plus importante que le biais d'agrégation introduit par l'IO dans l'ACVH.

2.3.2 Méthodologies de l'ACVH

Les données ACV et IO sont combinées dans un même inventaire. Toutefois, les jeux de données utilisés par l'ACV et l'IO sont souvent à un niveau de résolution différent. L'ACV étudie la production de bananes alors que l'IO s'intéresse aux fruits de manière générale. De plus, les données ACV et IO ne sont typiquement pas directement compatibles, car basées sur des unités différentes. L'ACV a tendance à prioriser les données physiques (e.g., kg, kWh, m^2, m^3, etc.) alors que l'IO opère typiquement en données monétaires. L'ACV étudie donc de 1kg de bananes alors que l'IO étudie 1€ de fruits. Combiner données ACV et IO requiert donc des concordances entre classifications des produits ainsi que des données de prix afin de pouvoir comparer les fonctions de production sur une même unité.

La littérature décrit quatre méthodes pour combiner données ACV et IO, i.e., quatre types d'ACVH : « tiered hybrid », « path exchange », « matrix augmentation » et « integrated hybrid » comme appelée dans Crawford et al. (2018). Chacune de ces méthodes consiste à combiner données ACV et IO. Cependant, certaines d'entre elles reposent davantage sur l'ACV ou au contraire reposent davantage sur l'IO.

Les quatre méthodes identifiées dans Crawford et al. (2018) proviennent de différentes initiatives (Bullard & Penner, 1978; Joshi, 1999; Lenzen & Crawford, 2009; Suh, 2004) et reposent sur des philosophies différentes. La communauté ACVH en revanche, semble en désaccord devant ces différentes méthodologies. Le cœur de l'article (Crawford et al., 2018) était de démontrer que la terminologie de l'ACVH et notamment de ses méthodologies est confuse. Les 95 études ACVH référencées dans l'article utilisent différentes dénominations pour les méthodologies de l'ACVH et cet article cherchait à poser une terminologie commune à l'ACVH et à présenter le fonctionnement de ces différentes méthodologies. Hélas, l'explication donnée sur le fonctionnement de chaque méthodologie vient, pour certaines, contredire la définition donnée dans l'article d'origine de ces méthodes.

Redéfinissons les buts, différences et similitudes de ces méthodologies :

La méthode par paliers (tiered hybrid dans son nom original) (Bullard & Penner, 1978) a pour but de compléter l'inventaire d'un processus élémentaire de l'ACV en y introduisant des données IO. On a donc des données d'avant-plan issues de l'ACV, des données d'arrière-plan également issues de l'ACV ainsi que des données d'arrière-plan IO venant compléter données ACV déjà présentes.

Par exemple, un procédé ACV de production de voiture n'incluant pas d'intrants de services (e.g., services informatiques) se verrait ajouter des intrants de services modélisés par l'IO. Les données ajoutées sont souvent appelées troncatures d'amont (upstream cut-off dans leur nom original, notées C^u) (Suh, 2004). La figure 2-7 donne une représentation de cette méthodologie. Un procédé de « production de voiture X » y est représenté. On y retrouve les matrices de l'ACV (comme illustrées en figure 2-2). La différence vient de l'introduction de la matrice de troncatures d'amont qui donne une estimation des intrants manquants, e.g. dans l'exemple de la figure 2-7, on estime que la production de voiture X manque 0.1€ de services bancaires. Le rôle de la matrice C^u est donc de contenir l'intensité des intrants manquants à ajouter pour ensuite lier ces intrants à leur modélisation dans l'IO (i.e. dans A^{io}).

Cette méthode respecte l'équilibre financier de l'IO (plusieurs techniques pour respecter cet équilibre sont détaillées par après en section 4.5.4), mais ne respecte pas l'équilibre des marchés. La philosophie de cette méthodologie est donc *d'obtenir un inventaire ACV spécifique et complet*.

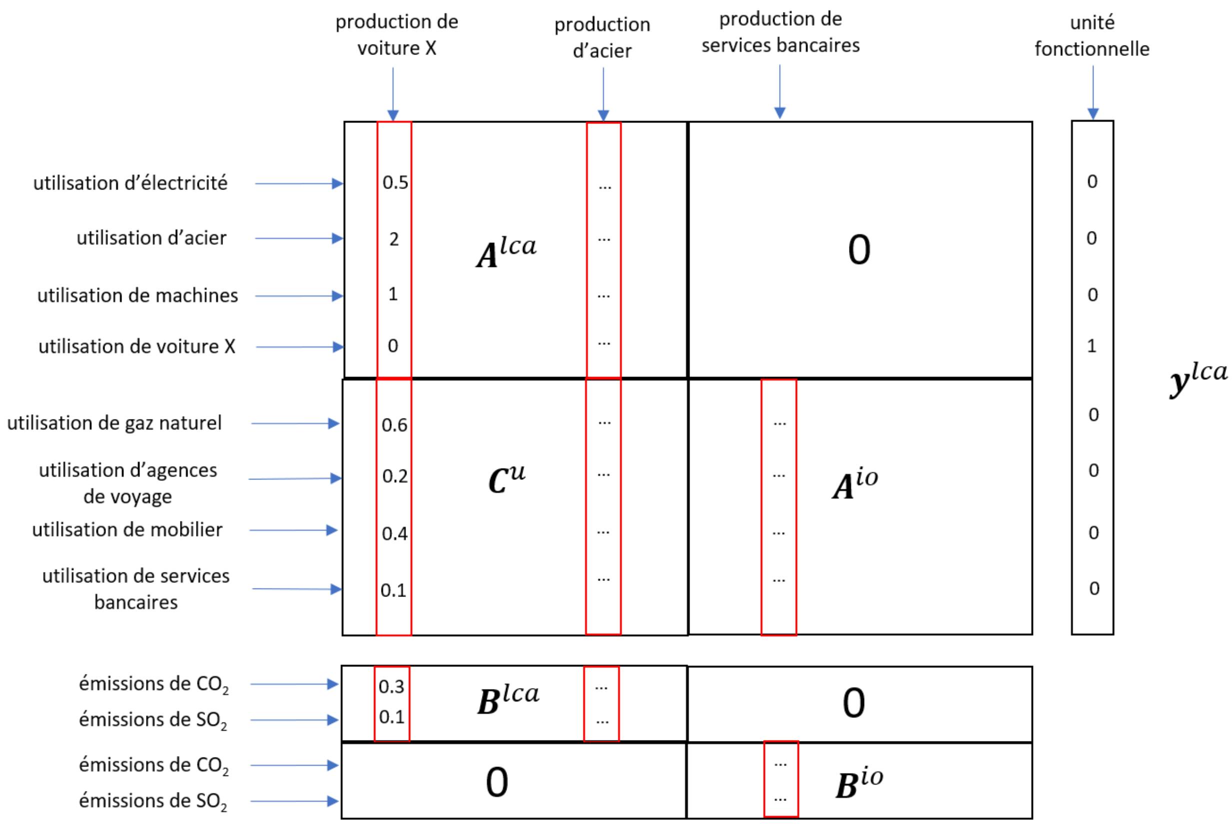


Figure 2-7 Représentation visuelle des matrices impliquées dans la methode par paliers

L'augmentation de matrice (matrix augmentation) (Joshi, 1999) et la méthode intégrée (integrated hybrid) (Suh, 2004) ont été considérées comme deux méthodologies différentes par Crawford et al. (2018). Cependant, toutes deux consistent à désagréger un secteur de l'IO en ayant recours à des données plus spécifiques, typiquement issues de l'ACV. Leur application mathématique est également identique. Ces deux méthodes sont en réalité la même. Par la suite, nous appellerons ces deux méthodes « méthode intégrée » uniquement. Pour désagréger un secteur de l'IO, les données ACV sont intégrées à la table IO, i.e., les secteurs de l'IO sont en mesure de s'approvisionner auprès des processus élémentaires intégrés plutôt qu'auprès de secteurs agrégés. Cela nécessite des troncatures d'aval (downstream cut-off dans leur nom original, notées C^d) décrivant l'achat par un secteur du produit issu d'un processus élémentaire. La figure 2-8 montre la représentation de la méthode intégrée. On note que les méthodes par paliers et intégrée partagent la même structure. En revanche, chaque donnée dans la rangée « utilisation de voiture X » décrit l'intensité de l'achat par chaque secteur (i.e. par chaque colonne de la matrice C^d) de voitures X. Cette information était déjà donnée pour les procédés ACV, mais est fixée à zéro dans la méthode par paliers (voir figure 2-7) pour les secteurs de l'IO. Le rôle de la matrice C^d est donc de contenir l'intensité des achats des procédés ACV par les secteurs de l'IO. Cela permet d'intégrer les procédés ACV dans l'économie représentée par l'IO.

Contrairement à la méthode par paliers, la méthode intégrée a pour point de départ l'IO. De plus, elle respecte à la fois l'équilibre financier et l'équilibre des marchés de l'IO. Afin de respecter l'équilibre financier, elle définit les mêmes troncatures d'amont que la méthode par paliers pour les processus ACV intégrés. Le but de la méthode intégrée est *d'apporter plus de spécificité à l'IO*. En d'autres termes, de permettre à l'IO de pouvoir étudier des produits de plus en plus précis (e.g., étudier l'impact de la production de bananes au lieu d'être limité à l'étude de la production de fruits et légumes). Les ACVH intégrées sont majoritairement appliquées pour réaliser des études prospectives sur les secteurs désagrégés. Par exemple, on pourrait utiliser une description ACV d'un véhicule électrique pour désagréger le secteur « véhicules motorisés » de l'IO et ainsi pouvoir étudier les effets d'une augmentation des ventes de véhicules électriques sur les autres secteurs de l'économie ainsi que sur l'environnement.

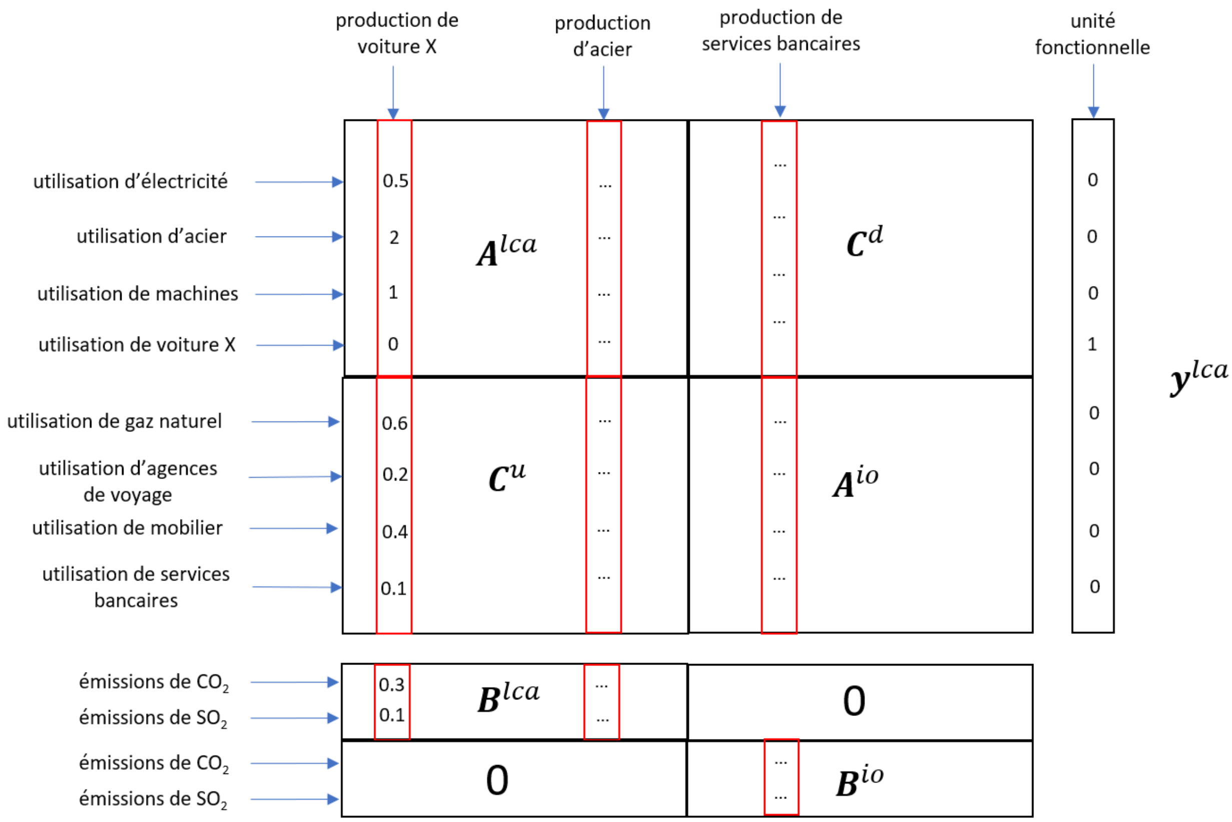


Figure 2-8 Représentation visuelle des matrices impliquées dans la methode intégrée

Finalement, le « path exchange » (Lenzen & Crawford, 2009) est une méthode à part qui ne fonctionne pas de façon matricielle comme les deux autres méthodes, mais fonctionne à travers l'analyse de chemin (structure path analysis) (Defourny & Thorbecke, 1984). Une matrice IO est découplée en différents chemins, e.g. de l'extraction de minerais jusqu'à la commercialisation de la voiture (voir figure 2-9). La méthode consiste ensuite à venir changer un nœud du chemin (i.e., une donnée précise dans le chemin, e.g., la quantité d'acier inox requise), mais de garder le reste du chemin inchangé. De ce fait, la chaine de valeur en amont de l'intrant n'est pas changée, contrairement aux méthodes matricielles. On introduit donc une donnée plus précise, ce qui vient apporter plus de spécificité. La philosophie de cette méthode consiste à *apporter une spécificité à l'échelle de l'intrant* et non de la chaine de valeur. Le « path exchange » pourrait cependant être appliqué de manière à apporter de la complétude à un inventaire ACV, en ajoutant des données manquantes au chemin créé à partir d'un procédé ACV.

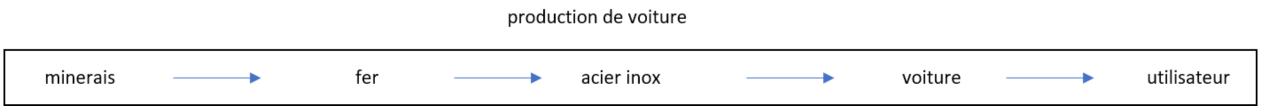


Figure 2-9 Représentation de la méthodologie « path exchange »

2.3.3 Deux approches distinctes pour la détermination du flux de données manquantes

Deux techniques peuvent être utilisées pour estimer la quantité d'intrants manquants à ajouter au processus élémentaire de l'ACV. Ces deux techniques ne sont pas clairement distinguées dans la littérature mais toutes deux existent. Adopter l'une ou l'autre a des conséquences toutefois, sur la quantité de données à collecter, sur la performance de l'inventaire obtenu et sur les actions subséquentes.

La première de ces techniques consiste à estimer les flux de données manquantes par une récolte de données plus poussée (Suh, 2004). La donnée qu'on sait absente est donc dans un premier temps estimée par récolte de données, puis modélisée avec l'IO. Par exemple, un procédé de fabrication de voitures de l'ACV n'inclut pas de services informatiques, ce qui constitue une troncation, car l'on sait que l'entreprise a besoin de services informatiques pour fonctionner. La quantité de services informatiques que l'entreprise utilise peut être obtenue auprès du service comptabilité de l'entreprise (e.g., 1000000€/an), puis allouée à chaque unité fonctionnelle (e.g., 2.50€/unité fonctionnelle). Cette quantité de services serait ensuite modélisée par l'IO, donnant la fonction de production pour 1€ de services informatiques moyen, résultant en des flux élémentaires et des impacts sur l'environnement supplémentaires. Cette technique est simple à mettre en œuvre. Elle est aussi précise, car il s'agit d'une donnée collectée et non estimée. Cependant, en plus de demander plus d'efforts de récolte de données au praticien ACV, elle requiert aussi que le praticien ACV soit conscient de toute troncature existante. Toujours avec le même exemple, le praticien pourrait ajouter une quantité de services informatiques (et donc réduire la troncature) mais pourrait ne pas penser à réduire la troncature sur d'autres intrants (e.g., services immobiliers, banquiers, etc.) ou pourrait également ne pas avoir accès à la donnée sur ces autres intrants. Ainsi, cette technique est précise, mais l'élimination de toute troncature existante n'est pas systématique.

La seconde technique consiste à s'appuyer sur la structure des données de l'IO afin estimer les quantités manquantes (Strømman, Peters, & Hertwich, 2009). *Tous* les intrants d'un secteur sont donc ajoutés, i.e., la structure des données de l'IO est ajoutée dans son ensemble. Ainsi, au lieu d'obtenir de la donnée supplémentaire (e.g., demander au service comptabilité combien de ressources informatiques ont été nécessaires), la quantité typiquement requise par les entreprises du secteur est utilisée. Plus concrètement, l'IO indique qu'en moyenne un producteur de voitures a besoin d'une certaine quantité de services informatiques par € de voiture fabriquée (e.g., 0.03€). En utilisant le prix de la voiture (e.g., 15000€), on ajouterait la quantité de service informatique correspondante (0.03 x 15000 = 450€). Cette technique ne requiert pas de récolte de données supplémentaires et ne requiert pas que le praticien connaisse toutes les troncatures possibles, car toutes les données de l'IO sont ajoutées, et donc toute troncature potentielle est retirée. Cependant, il y a nécessairement un double comptage qui s'opère avec cette technique (Strømman, 2009). En effet, les données déjà couvertes par l'ACV (e.g., l'aluminium de la voiture) sont comptées deux fois, puisque l'aluminium moyen utilisé dans la fabrication de voiture a également été ajouté (encore une fois, la structure de données complète de l'IO est ajoutée). Cette seconde technique doit donc constamment s'assurer que chaque intrant soit compté une seule fois, soit par l'ACV, soit par l'IO. On parle d'une étape de correction du double comptage. Cette étape n'est pas requise avec la première technique.

2.4 Limites de l'ACVH

2.4.1 Opérationnalisation de l'ACVH

Comme mentionné précédemment l'ACVH est rarement appliquée. La revue de littérature de Crawford et al. (2018) a uniquement identifié 95 études ACVH publiées entre les années 2010 et 2015. En effet, elle requiert davantage de données (notamment des données de prix), demande une plus grande expertise (il faut connaitre la méthodologie de l'ACVH et avoir des connaissances en ACV et en IO), n'est pas reconnue par ISO et aucun logiciel ne facilite son application. De plus, sa plus-value principale (la réduction de la troncature) n'a pas été démontrée de façon exhaustive. Il y a donc un travail d'opérationnalisation de l'ACVH à effectuer par la communauté afin de permettre aux praticiens ACV une application plus simple de l'ACVH. Il faut également obtenir une estimation représentative de la troncature en ACV pour encourager l'adoption de l'ACVH.

2.4.2 Un choix de méthodes de correction du double comptage aux effets inconnus

La correction du double comptage est une étape clé dans la création d'un inventaire hybride utilisant la seconde technique introduite dans la section 2.3.3. Plusieurs méthodes existent déjà (Strømman et al., 2009; Strømman & Solli, 2008; Wiedmann et al., 2011). Elles reposent sur des suppositions extrêmement différentes. Certaines considèrent que les données déjà présentes de l'ACV ne doivent pas être complétées par l'IO (Arvesen, Birkeland, & Hertwich, 2013) alors que d'autres méthodes au contraire considèrent que les données ACV déjà présentes devraient elles aussi être complétées par l'IO (Williams, 2004). Ce sont deux positions radicalement différentes qui auront un impact (majeur ou non) sur l'inventaire hybride obtenu. Et pourtant, ces méthodes ne font jamais l'objet de plus de quelques lignes dans un article scientifique. Il n'y a également aucune étude répertoriant et comparant les méthodes existantes. La communauté ACVH n'a donc pas d'idée du degré d'influence que le choix de telle ou telle méthode a sur les résultats obtenus.

2.4.3 Un arrière-plan non hybridé

Dans les quelques études ACVH existantes, seuls les procédés d'avant-plan sont hybridés. Ces procédés sont ensuite typiquement reliés à une base de données ACV (e.g., ecoinvent) pour les ACVH par paliers. Or, ces bases de données ne sont pas elles-mêmes hybridées. Chaque procédé de ces bases de données est donc lui-même tronqué. Dans ces études, le problème de troncature de l'ACV est donc réglé pour l'avant-plan, mais est toujours présent dans l'arrière-plan. L'ACVH appliquée de cette façon ne règle donc pas le problème de troncature dans son ensemble. La figure 5-1 du deuxième article de cette thèse illustre ce phénomène.

Une base de données hybride est donc requise pour complètement régler le problème de troncature de l'ACV, ce qui a déjà été encouragé à plusieurs reprises dans la littérature scientifique (Bontinck, Crawford, & Stephan, 2017; Crawford et al., 2018; Majeau-Bettez, Strømman, & Hertwich, 2011; Strømman et al., 2009). Plus concrètement, une base de données hybride est une base de données ACV où chaque procédé est hybridé, en utilisant une EEIO globale pour modéliser les données manquantes.

Plusieurs problèmes rendent l'obtention de telles bases de données compliquée. En effet, cela requiert l'hybridation semi-automatique de milliers de procédés, alors que l'hybridation est typiquement un processus manuel et nécessitant une compréhension poussée des flux et troncatures du procédé à hybrider. Il faut donc traduire la méthodologie en programme informatique non dépendant du procédé à hybrider, récolter des données de prix pour chacun des procédés, automatiser la correction du double comptage qui est majoritairement une étape manuelle, et ce pour des procédés de bases de données pour lesquels l'expertise concernant le procédé ne sera pas disponible.

2.4.4 Des capitaux absents de l'hybridation

Un capital (en IO) est un bien requis dans la production de produits ou services qui est utilisé pendant plus d'une année (European Commission, 2008). Puisque les bases de données IO sont tenues pour chaque année, les capitaux (qui durent plus qu'une année) sont traités différemment des autres intrants. Les investissements capitaux (e.g., l'achat d'un bâtiment) ne sont pas associés à la chaine de valeur du produit qu'ils aident à produire, mais sont plutôt associés à la demande finale, i.e., ils sont considérés au même titre que les dépenses du gouvernement et des consommateurs. Par conséquent, une voiture en IO ne requiert aucun intrant direct d'usine dans sa fonction de production, mais inclut une consommation du capital (e.g., maintenance du bâtiment).

Lors de l'hybridation, seuls les intrants directs de l'IO sont utilisés pour déterminer les intrants manquants de l'ACV. Donc, les capitaux n'étant pas intrants directs, ils ne sont typiquement pas ajoutés aux descriptions de l'ACV. Par conséquent, dans le cas où une donnée de capital serait manquante en ACV, elle ne pourrait pas être complétée par l'IO. Prenons le cas hypothétique dans lequel une étude ACV sur la production d'une voiture n'inclut pas d'usine, ce qui constitue une troncature. L'hybridation devrait donc venir combler ce manque de donnée, mais en est incapable, car l'usine de production de voiture n'est pas directement liée à la production de voitures en IO.

L'endogénéisation des capitaux est une pratique en IO consistant à venir lier les consommations de capitaux aux différentes chaines de valeur qu'ils aident à produire. Ainsi, ces chaines de valeur contiennent enfin des intrants directs de capitaux. Une version endogénéisée de bases de données IO voit son cadre méthodologique changer. L'ACVH a clairement intérêt à opérer avec des versions endogénéisées de l'IO mais doit donc adapter son propre cadre méthodologique en conséquence.

2.4.5 Une donnée de prix problématique

Comme introduit précédemment, toute ACVH requiert une donnée de prix afin de pouvoir combiner inventaires physique (ACV) et monétaire (IO). Typiquement, le prix du produit de référence, i.e., le produit étudié, est utilisé. Dans l'exemple de la section 2.3.3., le prix du produit de référence (la voiture) est utilisé pour estimer la quantité d'intrants manquants (service informatique). Que se passe-t-il, si le prix du produit de référence n'est pas disponible, ou si la donnée de prix n'est pas fiable ? L'hybridation n'est alors pas réalisable. Les troncatures de ce procédé ne peuvent alors pas être corrigées. De plus, chaque autre procédé utilisant ce procédé tronqué se retrouvera lui-même tronqué, de façon plus marginale.

En outre, l'hétérogénéité de prix de certains secteurs en IO a de grandes chances de mener à une estimation des données manquantes de l'ACV grandement sur ou sous-estimée. L'hétérogénéité dont on parle ici est celle du prix du produit de référence. Ce prix est celui utilisé afin de déterminer les quantités d'intrants manquants. Idéalement, il faudrait donc une bonne corrélation entre ce prix du produit de référence et la quantité d'intrants utilisés pour sa production, comme illustré en figure 2-10, afin d'obtenir la meilleure estimation possible de la quantité d'intrants manquants.

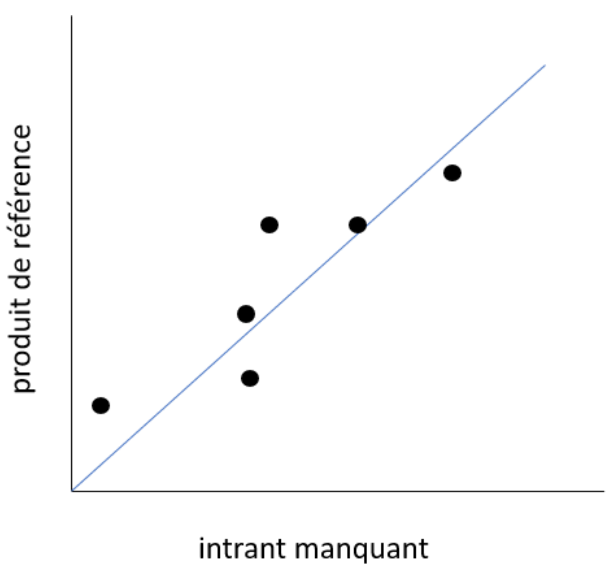


Figure 2-10 Une bonne corrélation entre le produit de référence et un intrant manquant

Dans un secteur avec une grande hétérogénéité de prix, la corrélation entre le prix du produit de référence et un intrant manquant quelconque a de grandes chances d'être mauvaise (figure 2-11). Alors, l'estimation de données manquantes basée sur cette corrélation sera de mauvaise qualité.

La dépendance à la donnée de prix du prix de référence en ACVH est donc problématique.

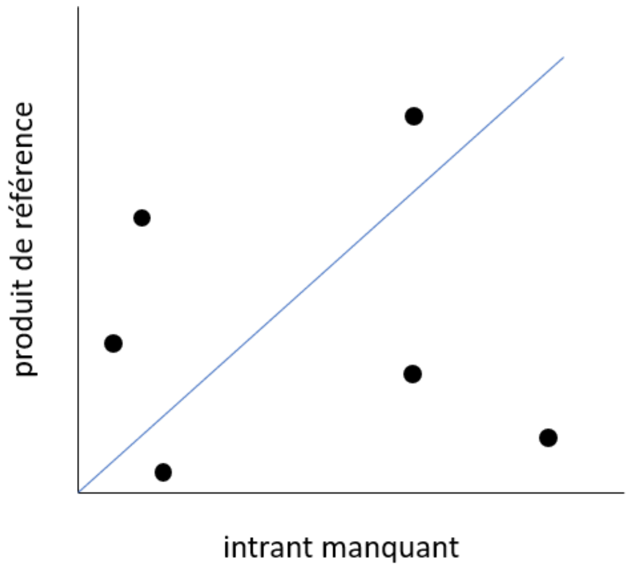


Figure 2-11 Pauvre corrélation entre le produit de référence et un intrant manquant.

2.4.6 Des émissions manquantes dans les EEIO globales

Le but premier de l'ACVH est de compléter l'inventaire de cycle de vie de l'ACV afin de réduire au mieux les troncatures existantes. La plupart des ACVH opèrent uniquement en termes de flux économiques, les impacts des données supplémentaires ajoutées par hybridation étant par la suite quantifiés par l'EEIO. Les flux élémentaires de l'ACV ne sont donc généralement pas modifiés eux-mêmes.

Cependant, les bases de données EEIO globales (e.g., EXIOBASE, GTAP, Eora, WIOD) ne contiennent que très peu de types d'émissions comparés aux bases de données ACV. EXIOBASE par exemple, ne contient que 35 types d'émissions alors qu'ecoinvent en contient plus de 700. Les impacts de l'IO sont donc susceptibles d'être sous-estimés à cause de la sous-représentation de certains types de polluants. L'ACVH, qui repose sur l'IO et ses extensions environnementales (ou flux élémentaires) pour modéliser les intrants manquants, souffre donc également de cette sous-estimation. Des extensions environnementales additionnelles seraient donc également nécessaires pour combler toutes les sous-estimations possibles de l'ACVH.

CHAPITRE 3 OBJECTIFS ET METHODOLOGIE GENERALE

3.1 Problématique

L'application de l'ACVH actuelle ne permet pas de corriger le problème de troncature de l'ACV, donnant uniquement lieu à une version un peu moins tronquée de l'ACV. La troncature peut prendre plusieurs formes comme exposé dans la revue de littérature qui tournent toutes néanmoins autour du manque de données, que ce soit au niveau de l'inventaire de l'ACV, de la non-endogénéisation des capitaux de l'IO, de la donnée de prix ou d'un nombre insuffisant d'émissions couvertes. Cette recherche s'intéresse donc à la question de recherche générale suivante : **comment améliorer la complétude des inventaires de cycle de vie tout en gardant la spécificité de l'ACV ?**

3.2 Définitions des objectifs de la thèse

L'objectif général de cette thèse est de **développer un cadre méthodologique permettant la génération de bases de données hybrides**. Cet objectif cherche à répondre aux questions de recherche spécifiques suivantes, venant des limites identifiées dans la section 2.4 :

- Quel est le niveau de troncature général de l'ACV ?
- Quel est l'impact des choix de méthodes de correction du double comptage sur les résultats obtenus ?
- Comment automatiser le processus d'hybridation ainsi que la correction pour le double comptage ?
- Comment opérer une hybridation avec une version endogénéisée d'une base de données EEIO ?
- Comment hybrider un procédé sans avoir nécessairement recours au prix du produit de référence ?
- Comment intégrer des extensions environnementales additionnelles au sein d'un cadre d'hybridation ?

Pour remplir l'objectif général, il faudra répondre aux différents objectifs spécifiques suivants :

- Développer une méthode de correction du double comptage semi-automatique
- Développer une base de données hybride pour la réalisation d'études comparatives ACVH
- Développer une méthode d'hybridation ne dépendant pas du prix du produit de référence
- Intégrer l'endogénéisation des capitaux et la génération de flux élémentaires additionnels dans le cadre d'hybridation développé

3.3 Méthodologie générale

La méthodologie générale suivie pour répondre à l'objectif de cette thèse est décrite dans la figure 3-1. La réponse aux quatre objectifs spécifiques décrits a donné lieu à l'écriture de trois articles scientifiques qui composent les 3 prochains chapitres de cette thèse.

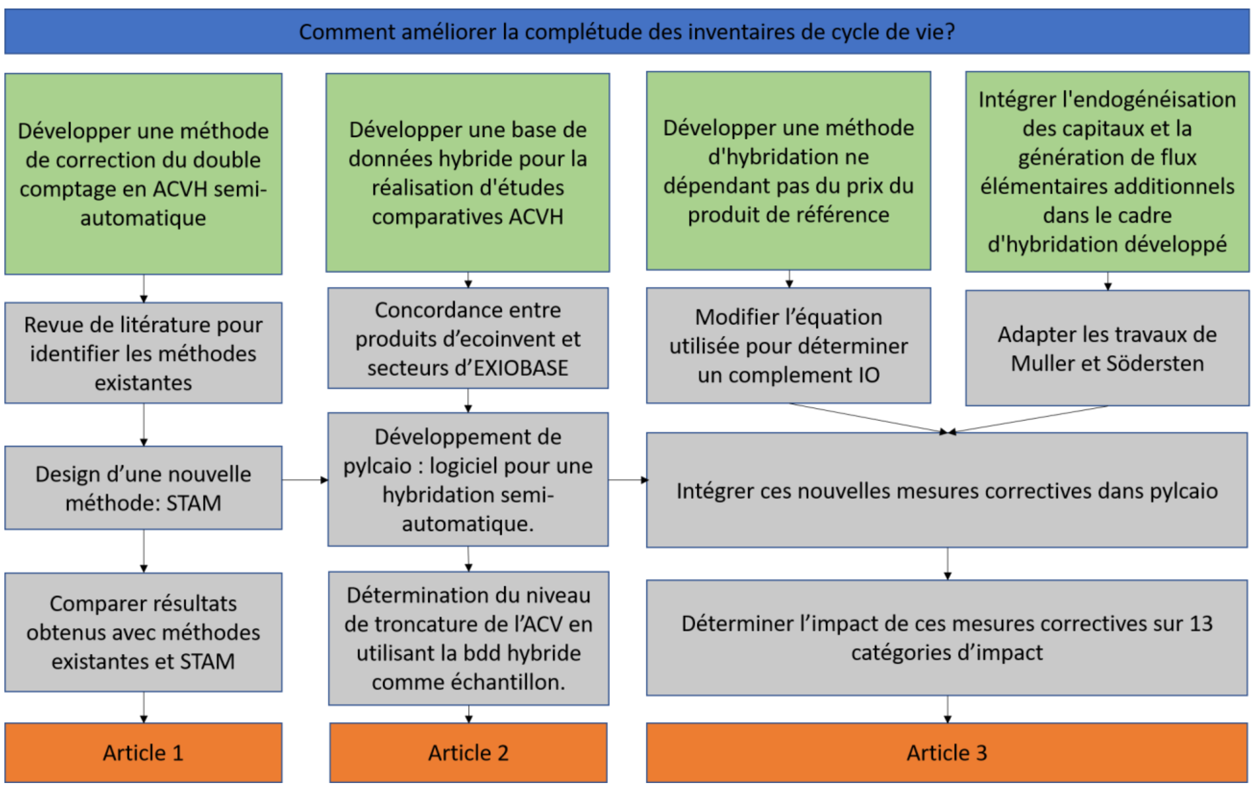


Figure 3-1 Méthodologie générale de cette thèse

3.4 Données et méthodologies sélectionnées

3.4.1 La base de données ACV sélectionnée

Afin d'appliquer le cadre d'hybridation, une base de données ACV est requise. Il en existe une multitude, pour en citer quelques-unes : ecoinvent (Wernet et al., 2016), GaBi (Gabi solutions, 2019), Agribalyse (ADEME, 2020), Agri-footprint (Blonk consultants, 2017), etc. Pour hybrider une base de données en revanche, certaines données/structure sont nécessaires. Premièrement, l'hybridation cherche en premier lieu à inclure toute l'économie. Aucune base de données ACV ne couvre la totalité de l'économie, mais deux d'entre elles couvrent une majorité de secteurs économiques : ecoinvent et GaBi. Il faut également des données de prix pour chacun des procédés, ce qui est disponible dans les versions 3+ d'ecoinvent. Enfin, il faut de la transparence et être capable de manipuler les données de la base de données (afin d'identifier les incidents de double comptage), ce qui n'est pas faisable avec GaBi. Pour toutes ces raisons, la base de données ecoinvent a été choisie. Plus spécifiquement, cette recherche utilise uniquement la version3.5 (Ecoinvent centre, 2018), car c'était la version la plus à jour au commencement de cette recherche et que la mise à jour à d'autres versions d'ecoinvent requiert un travail conséquent (concordance de classifications) qui n'a pas de lien avec la recherche.

3.4.2 La base de données IO sélectionnée

Pour hybrider la base de données ecoinvent, il faut également une base de données IO. Là encore plusieurs choix sont disponibles. La première distinction à faire est entre les tables IO nationales et multirégionales. Ecoinvent comprenant des chaines de valeurs globales, une table IO multirégionale et plus spécifiquement globale est requise. Quatre tables remplissent ce critère : EXIOBASE (Stadler et al., 2018), GTAP (Andrew & Peters, 2013), Eora (Lenzen, Moran, Kanemoto, & Geschke, 2013) et WIOD (Timmer, Dietzenbacher, Los, Stehrer, & de Vries, 2015). Les trois paramètres importants guidant le choix de la table IO à utiliser sont : le nombre de pays couverts, le nombre de secteurs économiques, le nombre de types d'émissions couvertes. Concernant le nombre de pays couverts et le nombre de secteurs économiques, Eora est mieux placée que les trois autres. En revanche, nombre de ces pays et secteurs économiques proviennent d'extrapolations ce qui diminue la qualité des données fournies par Eora. De plus, le nombre de secteurs couverts varie dépendamment des pays (e.g., 500 aux USA, mais 50 secteurs au Maroc), ce qui n'est pas souhaitable pour l'hybridation. EXIOBASE couvre 200 secteurs économiques sur 44 pays et 5 régions regroupant le reste des pays. Coïncidemment, les extensions environnementales d'EXIOBASE sont les plus développées. EXIOBASE a donc été choisie pour le cas d'application, plus spécifiquement la version 3.7 d'EXIOBASE (EXIOBASE Consortium, 2017).

Tableau 3-1 Tableau comparatif des différentes tables IO multirégionales globales

	Eora	EXIOBASE	WIOD	GTAP
Pays	187	44	41	129
Données extrapolées (%)	60	2	2	15
Secteurs	26-511	200	35	57
Extensions environnementales	++	+++	++	+

3.4.3 Endogénéisation des capitaux

La version d'EXIOBASE avec capitaux endogénéisés a été développée par Södersten et al. (2020) et sera utilisée dans cette recherche. C'est la seule version endogénéisée d'EXIOBASE disponible à l'heure actuelle.

3.4.4 Extensions environnementales additionnelles

Les extensions environnementales étendues pour EXIOBASE à partir de USEEIO (Yang, Ingwersen, Hawkins, Srocka, & Meyer, 2017a) produite par E. Muller (2019) seront utilisées dans cette recherche. Ces extensions ont été spécifiquement développées dans le but d'être intégrées dans le cadre d'hybridation présenté dans cette thèse.

3.4.5 Choix de la méthode ACVH

La méthodologie sélectionnée est la méthode par paliers telle que définie dans la section 2.3.2. Le but de cette recherche, i.e., le développement d'une hybridation automatisée afin de compléter les recettes fournies par l'ACV, est en parfait accord avec le but principal de la méthode par paliers. De plus, la structure du programme informatique pourra facilement être étendue pour passer de la méthode par paliers à la méthode intégrée.

3.4.6 Choix de la technique de génération de données en ACVH

Pour générer les estimations de données manquantes, ce projet de recherche se basera sur la structure de données fournies par l'IO. Le but de cette thèse, encore une fois, étant de développer une hybridation automatisée, la collecte de données supplémentaires n'a pas été considérée ici. Cela signifie que ce projet de recherche devra corriger pour le double comptage qui arrivera nécessairement, qui est d'ailleurs un point de recherche clé de cette thèse.

CHAPITRE 4 ARTICLE 1 : LIFTING THE VEIL ON THE CORRECTION OF DOUBLE COUNTING INCIDENTS IN HYBRID LIFE CYCLE ASSESSMENT

Ce chapitre est une reproduction de l'article « Lifting the veil on the correction of double counting incidents in hybrid life cycle assessment ». Cet article a été publié dans le *Journal of Industrial Ecology* (https://doi.org/10.1111/jiec.12945). Les co-auteurs de cet article sont : Guillaume Majeau-Bettez, Manuele Margni, Anders Hammer Strømman et Réjean Samson. Dans la suite de cette thèse, cet article sera référencé comme (Agez et al., 2019).

Summary

Life Cycle Assessment (LCA) and Environmentally Extended Input Output Analyses (EEIOA) are two techniques commonly used to assess environmental impacts of an activity/product. Their strengths and weaknesses are complementary, and they are thus regularly combined to obtain hybrid LCAs. A number of approaches in hybrid LCA exists, which leads to different results. One of the differences is how to make sure that LCA and EEIOA data to be mixed do not overlap, which is referred to as correction for double counting. This aspect of hybrid LCA is often ignored in reports of hybrid assessments and no comprehensive study has been carried out on it. This paper strives to list, compare and analyze the different existing methods for the correction of double counting. We first harmonize the definitions of the existing correcting methods and express them in a common notation, before introducing a streamlined variant. We then compare their respective assumptions and limitations. We discuss the loss of specific information regarding the studied activity/product and the loss of coherent financial representation caused by some of the correcting methods. This analysis clarifies which techniques are most applicable to different tasks, from hybridizing individual LCA processes to integrating complete databases. We finally conclude by giving recommendations for future hybrid analyses.

4.1 Introduction

Both Life Cycle Analysis (LCA) and Environmentally Extended Input Output Analyses (EEIOA) allow the quantification of emissions linked directly or indirectly to a product's production, use and end-of-life, which then enlightens on the environmentally-friendlier choice between competing products (Jeswani, Azapagic, Schepelmann, & Ritthoff, 2010). Even if their goal is the same, the methods do not operate in the same way. LCA aims at providing detailed descriptions of every process involved in the life cycle of a studied activity (International Organization for Standardization, 2006), while EEIOA regroups national inventories to describe the interdependence between economic sectors (R. E. Miller & Blair, 2009a).

The determination of the Life Cycle Inventory (LCI) according to ISO standards requires the description of all processes linked by the economic flows to the provision of a functional unit, along with all elementary flows that describe the exchanges of these processes with the environment (International Organization for Standardization, 2006). However, the economy is so deeply intertwined that it could be considered to be almost involved in its entirety in the life cycle of every single product (Gibon & Schaubroeck, 2017; Pomponi & Lenzen, 2017). Such broad coverage and inventory collection is currently impossible to achieve given the level of detail that is required in LCA and LCA practitioners strive to communicate the limited coverage of their studies by the definition of explicit system boundaries (Suh et al., 2004). System boundaries and coverage are not uniform and largely depend on data availability and research resources, complicating comparison between studies (Price & Kendall, 2012). Even harmonized databases struggle to ensure consistent coverage in process descriptions (Mongelli, Suh, & Huppes, 2005). In addition, the collection and use of data on services has always been a shortcoming in LCA, as can be shown by the underrepresentation of services in databases such as ecoinvent (Wernet et al., 2016). The data collection in LCA is therefore *never* complete since multiple inputs (deemed negligible or unavailable) are necessarily ignored. These missing inputs can either result from an omission of an input (e.g., no data available on the IT services required to craft a car) or from an underestimated requirement in the inventory (e.g., steel is accounted for in the inventory but is incomplete because bolts of steel were ignored). While an individual omission or underestimation may or may not impact the results of an LCA, the same cannot be ascertained for these omissions together which trigger what is referred to as truncations in LCA. Estimates of these truncations in the literature typically ranged from 20 to 50% (Ferrao & Nhambiu, 2009; Junnila, 2006; Lenzen, 2000; Lenzen & Dey, 2000; Norris, 2002; Rowley et al., 2009). The truncation of services alone is estimated to cause a systematic average underestimation of 6% (Ward, Wenz, Steckel, & Minx, 2017). Truncation errors in LCA are often being debated. Recently Yang *et al.* (2017) argued that truncation errors in LCA could be lower than aggregation errors in IO, which was disputed by Pomponi and Lenzen (2017). Another common argument against the problem of truncations in

LCA is that in a context of comparative studies, truncations between two similar products, which thus share similar value chains, should be similar themselves. Truncations would therefore not be a hindrance in comparative LCAs. While this is generally true, there are multiple cases where commodities that differ and rely on different value chains can fulfill a similar function, and in such cases LCA comparison results can switch, as proved in (Lenzen & Treloar, 2003).

EEIOA comes from the economic Input Output (IO) models, first developed by Leontief (Leontief, 1936) to capture indirect links between household consumption and employment and later extended to encompass waste flows and exchanges with the environment (Leontief, 1970). It is based on trading information that companies provide to their government, which result in national inventories (European Commission et al., 2009). The information is then compiled into a technology matrix (A) that gives average production functions (recipes) at sectoral level. For instance, the French industry of automobiles requires 0.0257€ of commodities from the French steel industry and 0.0024€ of commodities from the French electricity industry, etc. to produce 1€ of its products (Wood et al., 2015). These national input-output tables can be combined to obtain a global representation of the world economy as a Multi-Regional Input Output (MRIO) table (Andrew & Peters, 2013; Lenzen et al., 2013; Stadler et al., 2018; Tukker & Dietzenbacher, 2013). These tables being based on national accounts, they include every declared monetary transaction. As such, there are no problems of truncations for formal economic flows of an EEIOA (Lenzen, 2000); every value chain is captured in its entirety. It has to be noted though that informal economy is not captured by EEIOA (e.g., voluntary work, smuggling, self-employed people, etc.) and that satellite environmental emissions and resource extraction accounts may also have limited coverage. The downside of EEIOA is that the recipes are heavily aggregated, representing the average production of group of products rather than the manufacture of a specific product, notably to protect the secrets of individual companies. Consequently, an electric car and a diesel car can (depending on the IO table used) be included in a same product group (“Motor vehicles”) and would therefore have the same impact on the environment. It is therefore impossible to realize a study on a *specific* product using EEIOA alone (Suh et al., 2004).

LCA and EEIOA both have respective strengths and weaknesses: LCA can study specific products but suffers from truncations and inconsistent boundary definitions, whereas EEIOA has complete system boundaries (no truncations) but can only access information at the sectoral level. It is apparent that these techniques are complementary and they were thus combined into Hybrid Life

Cycle Analyses (HLCA) to obtain the best out of both methods. HLCA has been a focus of research for decades (Bullard & Penner, 1978; Joshi, 1999; Lenzen & Crawford, 2009; Moskowitz & Rowe, 1985; Nakamura & Kondo, 2002; Suh & Huppes, 2005).

By mixing LCA and EEIOA data to obtain a specific and complete inventory, HLCA requires to control that: (1) no data is *truncated*, meaning that all flows must be represented either by LCA or by EEIOA, (2) no data is *double counted*, meaning that all data of an inventory must be inventoried *only once* either by the LCA or by the EEIOA (i.e., a same input cannot be accounted for by both LCA and EEIOA at the same time). This control, which the hybrid community refers to as "correction for double counting" is an important part of the hybridization of an inventory. It is also complicated by the differences in aggregation level, measurement units (physical units used in LCA, monetary units used in EEIOA) and scope of LCA and EEIOA datasets. Many methods have been developed to tackle these issues, but these methods were created by isolated groups of scientists in the context of different studies with different purposes. It results in methods to correct double counting that require different amount of knowledge on the data used and that work with different assumptions, which ultimately leads to different results of hybridization. Despite being the source of differences in results, these methods have never been the focus of any study and as such, have never been systematically analyzed, compared, or even inventoried. Moreover, the hybrid community which was until recently solely focused on the hybridization of individual processes, is now looking towards massive hybridization, ultimately resulting in hybrid databases (Crawford et al., 2018; Stephan, Crawford, & Bontinck, 2018; Strømman et al., 2009; Suh & Lippiatt, 2012; Weidema, 2011). To achieve this overarching goal, the methods to correct double counting must be streamlined for application to a large number of processes about which less information is available (because the hybrid practitioner is not the one who collected the data).

4.2 Aim and scope

The present paper aims at reviewing and harmonizing the existing double counting correction techniques --- their assumptions, limitations, efficiency, applicability to database hybridization and effects on hybridization results --- and addressing some of their shortcomings through the development of two novel methods. All the methods will be introduced following the same example to allow direct comparisons and represented with matrices in a common notation (when possible) in order to better understand their functioning and differences.

The article first introduces the mathematical frameworks of LCA, EEIOA and HLCA with which we operate. We then introduce the example on which all methods will be applied. We continue by presenting the different double counting correction techniques and exposing their assumptions and limitations, resulting in a diagram that maps out the relationships between these models. We then discuss on inconsistencies brought about by some of the methods. Finally, we provide recommendations for future hybrid analyses on key data to include in the data collection process, as well as which correcting method to apply in which situation.

The article is focused on problems of double counting applied to hybridization of single products. It does not tackle issues related to double counting in the responsibility sharing between LCA processes (see (Gallego & Lenzen, 2005)) or concept of total flow (see (Szyrmer, 1992)). This article also does not introduce the different scopes of hybridization (tiered hybrid, matrix augmentation, integrated hybrid, waste IO and path exchange) but fully focuses on the methods to correct for double counting (see (Crawford et al., 2018) and (Suh & Huppes, 2005)). The matrix notation and the structure of the hybrid technology matrix employed throughout this article are not novel and do not constitute an original contribution of this study (Perkins & Suh, 2019; Strømman et al., 2009). Our analysis of the double counting correction techniques is generally applicable to all hybridization scopes, because they all share the common challenge of harmonizing the boundaries between LCA and EEIOA inventories, as demonstrated in SI-1. This study focuses on the use of EEIOA as complement to LCA process data (i.e., upstream cut-offs) and does not address the disaggregation of EEIOA product groups with LCA data (i.e., downstream cut-offs). Finally, this study only concerns HLCAs relying on EEIOA to estimate missing inputs to processes , see section "*An important distinction between two different uses of EEIOA data in hybrid analyses*".

4.3 Hybrid LCA

4.3.1 An important distinction between two different uses of EEIOA data in hybrid analyses

To clarify how and why the need to correct for instances of double-counting arises, we must first distinguish between two conceptually distinct uses of EEIOA data in hybrid LCAs, each solving a particular problem. In other words, while all hybrid LCAs rely on EEIOA as a complementary source of data, the types of data gaps that are filled by this complement differ across studies.

In some HLCAs, EEIOA tables are used more or less like any other generic ("background") LCA database: they are used to model the flows upstream in the value chain of an inventoried (i.e., known) product flow of the system under investigation. Thus, a practitioner collects primary data to inventory the flows of a given ("foreground") unit process, and some of these flows may then be modelled as provided by another LCA process (e.g., from ecoinvent), or they may be modelled as being provided by a corresponding EEIOA sector. For example, in figure 1a, the practitioner identifies input flows of steel, electricity, and IT services. While the production of the first two are readily described by LCA processes (boxes in light grey), the delivery of IT services may prove difficult to inventory and may thus be unavailable in LCA databases. Relying instead on an average IO sector description for the provision of IT services allows the analyst to describe the value chain of this known input flow (boxes in dark grey). In short, since all sectors of the economy are present in an EEIOA database, an approximate upstream value chain description can always be found for any inventoried input flow, and these flows no longer need to be excluded from the system boundary due to lack of upstream data, as exemplified in (Fernández-Dacosta et al., 2017)[2], (Gibon et al., 2015)[3], (Suh, 2004)[4].

In contrast, some other HLCAs rely on EEIOA sector descriptions to estimate any potentially remaining missing direct inputs flows of an LCA process. In the example figure 1b, if the LCA practitioner is not confident that the three inventoried inputs (in red) constitute the totality of the requirements of the process, any missing inputs, e.g., wholesale trade, research and development, hotel services and so on (fig. 1b, in blue), may be estimated based on the requirements of similar

[2] "The 2011 dataset from the EXIOBASE 3.3 environmentally extended, multi-regional supply-use/input-output database was used to model the economic background for infrastructure of some SAs for hybrid modelling"

[3] “In a hybrid LCA, the foreground system typically requires both physical inputs from the process LCI database and economic inputs from the input−output database.”

[4] “The upstream cut-off by processes matrix is derived by dividing the total bill of goods for the inputs that are not covered by a processes in a process-based system during the period of steady-state approximation by the total unit operation time of each process.”

processes aggregated in the corresponding EEIOA sector. (Arvesen & Hertwich, 2011)[5], (Perkins & Suh, 2019)[6], (Strømman et al., 2009)[7] all explicitly rely on this strategy.

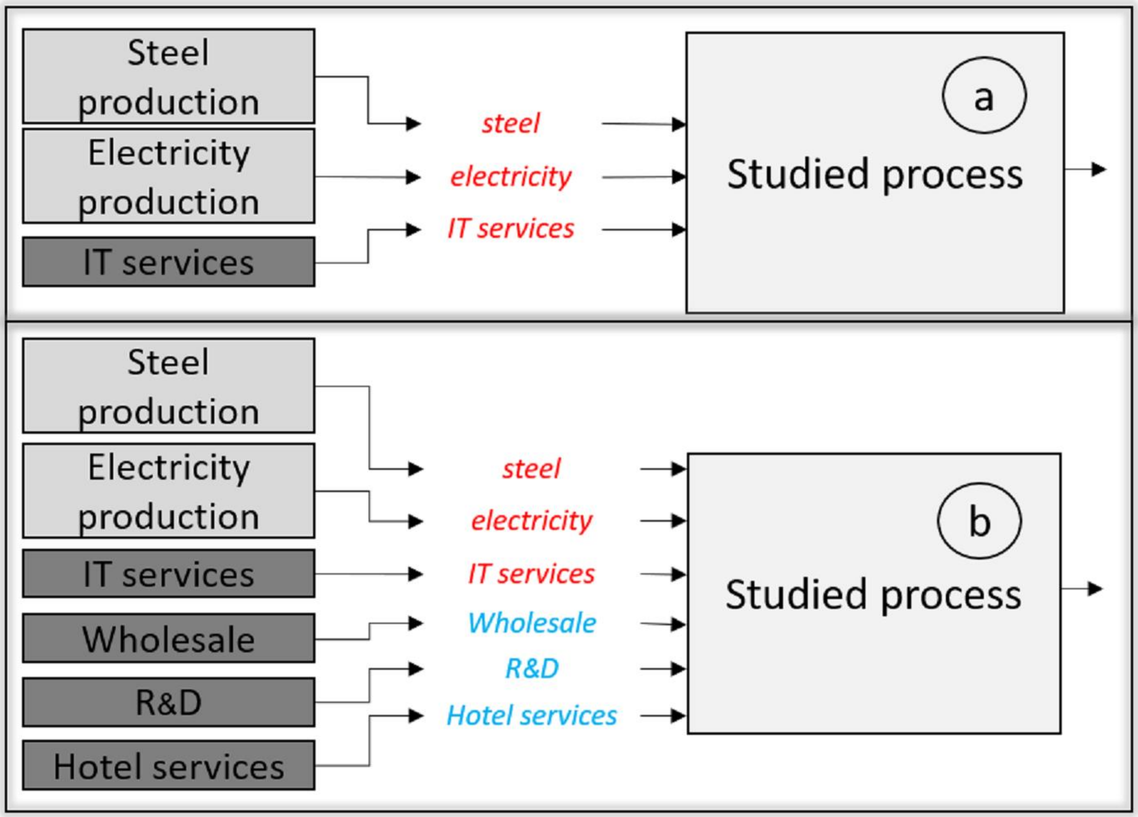


Figure 4-1 Example showing two ways EEIOA data can be used in hybrid analyses. In a, the practitioner inventoried three inputs (in red), of which one is not inventoried in general LCA database (boxes in light grey) and is thus represented by IO (boxes in dark grey). The sum of the cost of these three inputs + value-added amounts to the production cost of the studied process. In b, that sum is smaller than the production cost. There are thus missing inputs, which are (in this case) estimated (in blue) based on the IO data structure of the corresponding sector.

---

[5] "The foreground processes are assigned the same input distributions as their belonging IO sector"

[6] "[...] the "management of companies and enterprises" commodity was not already included in the process LCA and so was added *using input-output data* during the hybridization."

[7] "To achieve a more robust inventory compilation procedure we require *all* the available data on monetary inputs"

The first use is especially appropriate if the practitioner is confident that he or she had access to all direct inputs to the LCA process, notably if the sum of the cost of these inputs and value-added leads to complete cost coverage. Otherwise, there is a high risk that some input flows have been ignored, and that environmental burdens are still underestimated. The second use strives to fill this remaining data gap to correct for this negative bias. In doing this, care must be taken to avoid any overlap between the primary data and these new highly aggregated input estimates, which would lead to a positive bias, hence the need to correct for double-counting.

As this need to ensure no overlap between LCA and EEIOA only arises with the second usage, it will be the sole focus of our article. Throughout the following analysis, it should be kept in mind that the attempt is to estimate missing direct input flows of a process (and associated lifecycle requirements) where further primary data collection proved infeasible.

Of course, the above distinction between the two uses of EEIOA data is purely conceptual. In practice, these can be combined. Furthermore, although we have never encountered this in the literature, it would be possible to use EEIOA tables to only determine the quantity of the missing flows (blue inputs) without relying on these tables to represent these flows' upstream value chains. The upstream value chains of these additional flows could then be modeled through other data sources (e.g., the LCA database itself, if an appropriate process is available). Here, modeling upstream value chains with EEIOA or another source has no implication on the focus of this article, i.e., on the correction for the double counting, refer to SI-1.

4.3.2 What is double counting?

Hybridization mixes LCA and EEIOA system descriptions to obtain specificity and completeness. Unit process inventories from LCA are therefore completed with IO complements, i.e. an estimate of what was deemed missing in the original LCA inventory (see equation 1).

$$\text{all inputs} \;=\; \text{LCA inputs} \;+\underbrace{\text{estimates of missing inputs}}_{IO\ complements} \tag{1}$$

When estimating the missing inputs of a given LCA process using EEIOA, we (1) identify the equivalent EEIOA production function and all its inputs, but (2) we only keep the inputs that do not overlap with the flows already inventoried in the foreground. This means that, for the sake of this analysis, the hybridization described in equation 1 can be conceptually split into two formal steps. First, all the requirements of the corresponding EEIOA production function are added to the incomplete LCA production function we wish to hybridize. This step complements the missing inputs, but necessarily (and explicitly) represents twice the inputs originally present in the LCA inventory. As a second conceptual step, these excessive inputs are removed to avoid double-counting, hence the name "correction for double counting" in the literature (Strømman et al., 2009) (see equation 2).

$$\text{all inputs} = \underbrace{(\text{LCA inputs} + \text{all EEIOA inputs})}_{uncorrected\ hybridization} - \underbrace{\text{overlapping inputs}}_{correction} \qquad (2)$$

By definition, these two conceptual substeps can equivalently be performed in one step, where EEIOA inputs are corrected and added to LCA inputs, but, for clarity, and because this article focuses on this second conceptual step, we will systematically present them as distinct. To explore correction for double counting, we first show the double counting, and then we show the (various forms of) correction. For example, the hybridization of a process 'production of a car' will include after the first step, inputs of services and of steel from the average production of a vehicle according to EEIOA. While the introduction of services data is welcome because the original inventory poorly considered them, inputs such as steel were already included in the original inventory. The inputs of steel are thus accounted for twice in the resulting hybrid inventory, which leads to an overestimation if left uncorrected.

4.3.3 Mathematical framework

The results of hybridization will be described using matrices. As such, we need to introduce LCA's, EEIOA's and HLCA's mathematical frameworks. LCA and EEIOA use different conventions as to how to represent their respective technology matrices. Throughout the article we chose to use

EEIOA's convention, but at any point in the article one can switch to LCA's convention[8]. See table 1 for a summary of the different notations of this article.

Table 4-1 Notations used throughout this article

Symbol	***Coefficient***	***Description***
$\boldsymbol{m, n, k}$		LCA products or processes (e.g., pump, stainless steel/production of pump, production of stainless steel)
$\boldsymbol{M, N, K}$		EEIO product groups (e.g., Machinery, Steel)
$\mathcal{M}, \mathcal{K}$		STAM categories of product groups which share similar <u>physical descriptions</u> (e.g., Machines, Metals)
$\boldsymbol{\mathcal{F}}$		STAM categories of product groups which share similar <u>functionalities</u> (e.g., Liquid Fuels)
$\boldsymbol{I}$		Identity matrix
$\boldsymbol{A^{lca}}$	a_{mk}^{lca}	LCA technology matrix
$\boldsymbol{B^{lca}}$	b_{mk}^{lca}	LCA environmental extensions matrix
$\boldsymbol{y^{lca}}$	y_{k}^{lca}	LCA functional unit vector
$\boldsymbol{q^{lca}}$	q_{k}^{lca}	LCA vector of total environmental interventions

[8] To pass from EEIOA to LCA notations, use the relationship $\boldsymbol{T} = (\boldsymbol{I} - \boldsymbol{A})$ where $\boldsymbol{T}$ corresponds to the technology matrix using LCA's convention, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{A}$ corresponds to the technology matrix from EEIOA's convention.

Table 4-1 Notations used throughout this article (suite)

$\boldsymbol{A^{io}}$	$\boldsymbol{a_{MK}^{io}}$	EEIO technology matrix
$\boldsymbol{B^{io}}$	b_{MK}^{io}	EEIO environmental extensions matrix
$\boldsymbol{y^{io}}$	y_K^{io}	EEIO final demand vector
$\boldsymbol{q^{io}}$	q_K^{io}	EEIO vector of total environmental interventions
$\boldsymbol{q^{hyb}}$	q_K^{hyb}	Hybrid vector of total environmental interventions
$\boldsymbol{C^u}$		Upstream complement matrix
	$c_{Mk}^{uncorrected}$	Uncorrected coefficient of the upstream complement matrix
	c_{Mk}^{binary}	Coefficient of the upstream complement matrix corrected using the binary method
	c_{Mk}^{SSM}	Coefficient of the upstream complement matrix corrected using the SSM
	c_{Mk}^{BPC}	Coefficient of the upstream complement matrix corrected using the BPC
	c_{Mk}^{STAM}	Coefficient of the upstream complement matrix corrected using the STAM
$\boldsymbol{C^d}$	c_{mK}^{d}	Downstream complement matrix
$\boldsymbol{H}$	h_{Kk}	Concordance matrix product group/product
$\mathbf{G}$	$g_{M\mathcal{M}}$	Concordance matrix category/product group
$\boldsymbol{\pi}$	π_k	Vector of prices for LCA products

Table 4-1 Notations used throughout this article (suite)

$\boldsymbol{\Lambda}$	$\boldsymbol{\lambda_{Mk}}$	Binary filter matrix
$\boldsymbol{\Theta}$	θ_{Mk}	SSM filter matrix
$\boldsymbol{\Gamma}$	$\gamma_{\mathcal{MK}}$	STAM filter matrix
$\boldsymbol{\Phi}$	φ_{Mk}	STAM filter matrix
$= \mathbf{0}^{*}$		Real zero in the inventory (i.e., zero coming from the specificity of the inventory and not because of the unavailability of data)

LCA's mathematical framework has been formalized by Heijungs and Suh (2002) and is (using EEIOA's convention) as follows:

$$\boldsymbol{q}^{lca} = \boldsymbol{B}^{lca} \cdot (\boldsymbol{I} - \boldsymbol{A}^{lca})^{-1} \cdot \boldsymbol{y}^{lca} \tag{3}$$

where vector $\boldsymbol{y}^{lca}$ contains the functional unit, matrix $\boldsymbol{A}^{lca}$ is the technological matrix defining the amount of technological inputs per process, matrix $\boldsymbol{B}^{lca}$contains the different environmental interventions per process and vector $\boldsymbol{q}^{lca}$ quantifies the total environmental interventions linked to the functional unit over the whole life cycle.

EEIOA is based on the famous Leontief's equation (Leontief, 1936) from Input Output:

$$\boldsymbol{x} = \left(\boldsymbol{I} - \boldsymbol{A}^{io}\right)^{-1} \cdot \boldsymbol{y}^{io} \tag{4}$$

in which $\boldsymbol{x}$ is the total output of the economy, $\boldsymbol{A}^{io}$ the technological matrix and $\boldsymbol{y}^{io}$ the final demand.

Environmental extensions have then been included to transform equation (4) as follows (Leontief, 1970):

$$\boldsymbol{q}^{io} = \boldsymbol{B}^{io} \cdot \left(\boldsymbol{I} - \boldsymbol{A}^{io}\right)^{-1} \cdot \boldsymbol{y}^{io} \tag{5}$$

in which $\boldsymbol{y}^{io}$ is the final demand, $\boldsymbol{A}^{io}$ is the technology matrix, $\boldsymbol{B}^{io}$ is the matrix of the environmental burdens (thus "environmentally extended") and $\boldsymbol{q}^{io}$ corresponds to the total emissions linked to the final demand.

It results that hybrid LCA follows the next equation (6) (Crawford et al., 2018):

$$\boldsymbol{q}^{hyb} = [\,\boldsymbol{B}^{lca}\ \boldsymbol{B}^{io}] \cdot \left(\boldsymbol{I} - \begin{bmatrix} \boldsymbol{A}^{lca} & \boldsymbol{C}^{d} \\ \boldsymbol{C}^{u} & \boldsymbol{A}^{io} \end{bmatrix}\right)^{-1} \cdot \begin{pmatrix} \boldsymbol{y}^{lca} \\ 0 \end{pmatrix} \tag{6}$$

where $\boldsymbol{y}^{lca}$ corresponds to the functional unit of the system to hybridize, $\boldsymbol{A}^{lca}$ and $\boldsymbol{A}^{io}$ are the technology matrices for LCA and EEIOA respectively, $\boldsymbol{C}^{u}$ contains the upstream IO complements (often also called upstream cut-off matrix (Suh, 2004)), $\boldsymbol{C}^{d}$ corresponds to the downstream complements (i.e., the use of LCA inventoried commodities within the rest of the economy; often set to zero), $\boldsymbol{B}^{lca}$ and $\boldsymbol{B}^{io}$ represent the environmental interventions linked to the LCA and EEIOA parts of the hybrid system and $\boldsymbol{q}^{hyb}$ gives the total emissions linked to the hybrid inventory.

The determination of the inputs of $\boldsymbol{C}^{u}$ is not defined in an equation to the best of the authors knowledge. This determination was extensively described in the literature however (Strømman et al., 2009; Strømman & Solli, 2008; Yu & Wiedmann, 2018) and can accordingly be translated into the following equation (where $c^{uncorrected}$ are coefficients of $\boldsymbol{C}^{u}$ prior to correcting for double-counting):

$$c_{Mk}^{uncorrected} = a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k \tag{7}$$

The EEIOA requirement of product group M for the production of product group K (a_{MK}^{io}) is multiplied by the coefficient of the concordance matrix matching LCA process k to its corresponding product group K (h_{Kk}) and by the price of product k (π_k). This gives the uncorrected upstream cut-off coefficient ($c_{Mk}^{uncorrected}$), complementing process k with data on product group M, which has to be corrected to avoid potential overestimation. Refer to SI-2 for a demonstration of the determination of uncorrected upstream cut-offs.

4.4 Correction for double counting

4.4.1 Introduction of the example

The methods to correct for double counting will be introduced through an example: the production of a pump. The simplified production of this fictional pump relies on 0.2kg of stainless steel and 0.8kg of pre-fabricated parts composed of the same stainless steel. Let it be a given that the production of this pump does not require any kind of plastics and that the process description is incomplete thus requiring to be complemented through hybridization (figure 2 a)).

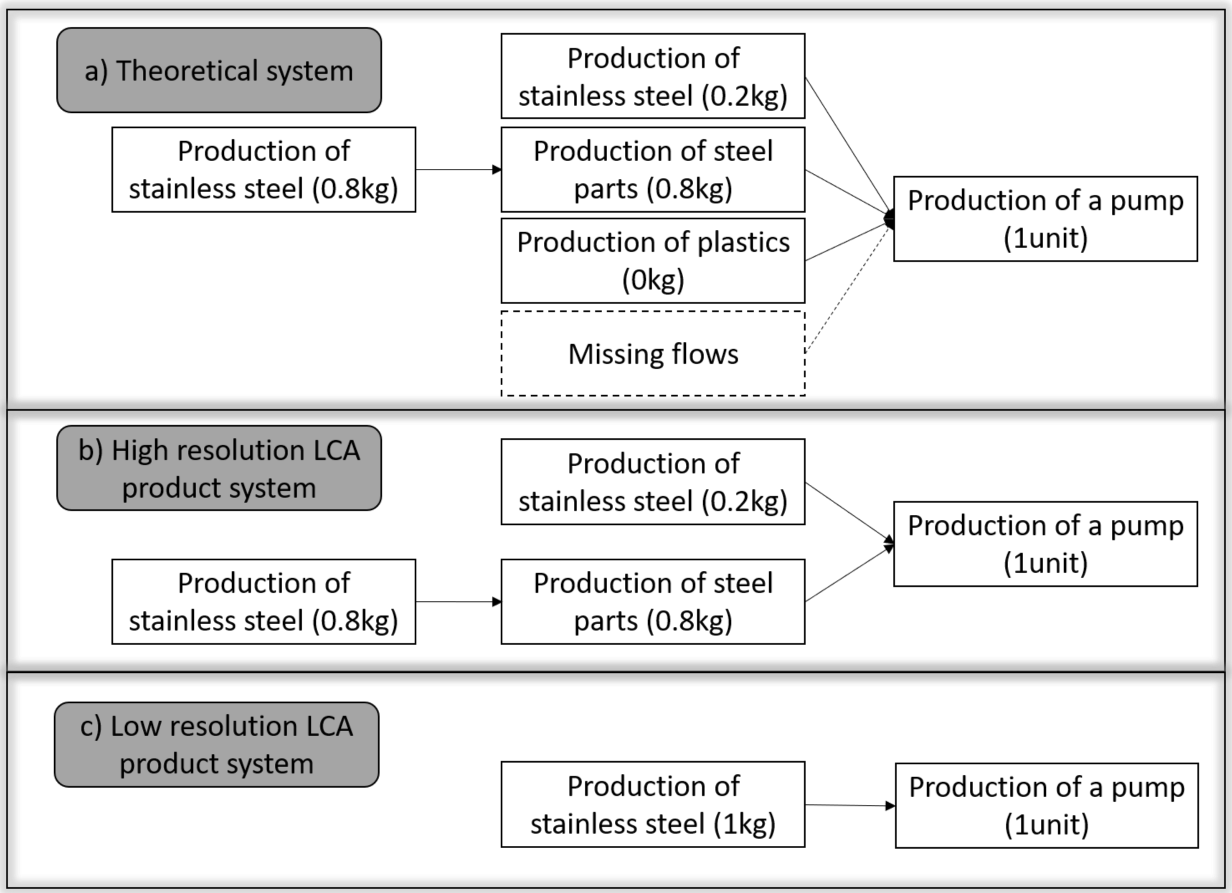


Figure 4-2 A same system can be represented with different resolutions in product systems and lose information

Figure 2 b) shows how LCA would represent the production of the fictional pump. Missing flows are obviously not explicitly in the product system and the null inputs are typically not entered explicitly either. As a result, the information according to which the production of the pump does not require any inputs of plastics is not seen in an LCA product system. Moreover, the fact that the parts are pre-fabricated is often either unknown or ignored depending on the resolution. Using a low-resolution product system can result in an aggregation of steel and steel parts into steel only (figure 2 c)). Despite the difference, both systems represented in figure 2b and 2c would yield identical LCA estimates of exchanges with the environment. In order to expose some of the shortcomings of the different methods to correct for double counting, we will consider having a low-resolution product system (figure 2 c)).

In figure 3, the hybrid technology matrix is represented, with $\boldsymbol{A}^{lca}$ containing the LCA data on the production of the pump, $\boldsymbol{A}^{io}$ is the commodity-by-commodity technology matrix with EEIOA data (inspired from Exiobase (Wood et al., 2015)), $\boldsymbol{C}^{u}$ and $\boldsymbol{C}^{d}$ are empty for now because the system is not yet hybridized. The production of stainless steel and plastics have no impact on the hybridization of the pump, and their values are therefore omitted and simply denoted by “n” for number.

The inventory of the production of the pump is truncated since there are missing or incomplete inputs, we thus hybridize this process. We assess that it belongs to the Machinery product group (through a concordance matrix $\boldsymbol{H}$) and the fictional price of the pump ($\pi = 12$€/unit) is then used to convert the EEIOA data (from Machinery) into $\boldsymbol{C}^{u}$ coefficients. Figure 4 represents this first step, thus resulting in an uncorrected hybrid system. Notice that the steel constituting the pump is inventoried both by LCA and EEIOA: it is therefore double counted which thus requires correction.

	Production of									
	Pump (unit)	Stainless steel (kg)	Plastics (kg)	Machinery (€)	Steel (€)	Fabricated metal (€)	Plastic (€)	Diesel (€)	Lubricant (€)	Other (€)
Pump (unit)	0	n	n	0	0	0	0	0	0	0
Stainless steel (kg)	1	n	n	0	0	0	0	0	0	0
Plastics (kg)	0	n	n	0	0	0	0	0	0	0
Machinery (€)	0	n	n	0.07	0.02	0.01	0.01	0.003	0.003	0.01
Steel (€)	0	n	n	0.03	0.2	0.1	0	0	0	0.006
Fabricated metal (€)	0	n	n	0.07	0.02	0.085	0.01	0.004	0.004	0.01
Plastic (€)	0	n	n	0.001	0	0.001	0.001	0	0	0.01
Diesel (€)	0	n	n	0.001	0	0	0	0	0	0
Lubricant (€)	0	n	n	0.001	0.002	0	0	0	0	0
Other (€)	0	n	n	0.4003	0.4755	0.5583	0.7922	0.7022	0.7098	0.621
Added value (€)	0	n	n	0.4577	0.2825	0.2457	0.1868	0.2908	0.2832	0.343
CO_2 (kg)	0.1	n	n	0.064	1.651	0.073	0.063	1.018	0.984	0.356

A^{lca} C^{d} C^{u} A^{io}

Figure 4-3 Representation of the hybrid system, pre-hybridization (A^{lca}=LCA technology matrix, A^{io}=EEIOA commodity-by-commodity technology matrix, C^{u}=upstream complements, C^{d}=downstream complements)

	Production of									
	Pump (unit)	Stainless steel (kg)	Plastics (kg)	Machinery (€)	Steel (€)	Fabricated metal (€)	Plastic (€)	Diesel (€)	Lubricant (€)	Other (€)
Pump (unit)	0	n	n	0	0	0	0	0	0	0
Stainless steel (kg)	1	n	n	0	0	0	0	0	0	0
Plastics (kg)	0	n	n	0	0	0	0	0	0	0
Machinery (€)	0.84	n	n	0.07	0.02	0.01	0.01	0.003	0.003	0.01
Steel (€)	0.36	n	n	0.03	0.2	0.1	0	0	0	0.006
Fabricated metal (€)	0.84	n	n	0.07	0.02	0.085	0.01	0.004	0.004	0.01
Plastic (€)	0.012			0.001	0	0.001	0.001	0	0	0.01
Diesel (€)	0.012			0.001	0	0	0	0	0	0
Lubricant (€)	0.012	n	n	0.001	0.002	0	0	0	0	0
Other (€)	4.8036	n	n	0.4003	0.4755	0.5583	0.7922	0.7022	0.7098	0.621
Added value (€)	5.1324	n	n	0.4577	0.2825	0.2457	0.1868	0.2908	0.2832	0.343
CO_2 (kg)	0.868	n	n	0.064	1.651	0.073	0.063	1.018	0.984	0.356

$Price_{pump}$ = 12€/unit

Figure 4-4 First step of the hybridization process: conversion of corresponding EEIOA data using the price of the product (requires correction for double counting)

4.4.2 Existing methods to correct for double counting

4.4.2.1 Binary method

The first identified method has been regularly used in several hybrid LCAs without ever being named (Arvesen et al., 2013; Arvesen & Hertwich, 2011; Deng, Babbitt, & Williams, 2011; Lin, Liu, Meng, Cui, & Xu, 2013; Stephan et al., 2018; Wiedmann et al., 2011). Given its nature, we chose to refer to it as the *binary method*. The method consists in inserting a matrix ($\boldsymbol{\Lambda}$) into equation (7), containing information on which inputs are explicitly covered by LCA or not, essentially filtering which EEIOA inputs to set to zero in the uncorrected upstream cut-off matrix, resulting in the (binary) corrected upstream complement matrix ($\boldsymbol{C}^u$), whose coefficients are determined following equation (8).

$$c_{Mk}^{binary} = \lambda_{Mk} \cdot a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k \tag{8}$$

$$c_{Mk}^{binary} = \lambda_{Mk} \cdot c_{Mk}^{uncorrected} \tag{8'}$$

The determination of the coefficients of $\boldsymbol{\Lambda}$ works as follows: whenever any input (m) is explicitly included into the process-based inventory ($a_{mk}^{lca} \neq 0$), the coefficient of its corresponding product group (M, where $m \in M$) will be forced to zero ($\lambda_{Mk} = 0$). As more than one LCA product can be matched to an EEIOA product group, a summation term is included in equation (9) to ensure that the EEIOA complement is set to zero whenever *any* m element of M is found in the LCA inventory.

$$\lambda_{Mk} = \begin{cases} 0 \; if \sum\limits_{m \in M} a_{mk}^{lca} \neq 0 \\ 1 \; if \sum\limits_{m \in M} a_{mk}^{lca} = 0 \end{cases} ,\{\forall M, k \tag{9}$$

In our example, the input coefficient from the steel product group (M) is forced to zero, because the process-based inventory already explicitly includes 1kg of stainless steel (m); all the EEIOA complements from other product groups are kept (figure 5 a)).

4.4.2.2 Strømman and Solli's method

The second identified method was developed by Strømman and Solli (2008) but was not named. Strømman & Solli's method (SSM hereafter) is similar to the binary method, in the sense that it also forces to zero all of the inputs from product groups whose products are explicitly included in the process-based inventory. The difference between the two methods comes from the introduction of another matrix (called theta) which allows the practitioner to force EEIOA inputs to zero for data that are known (or considered) to be complete, including null inputs. Equation (7) is therefore modified as:

$$c_{Mk}^{SSM} = \theta_{Mk} \cdot \lambda_{Mk} \cdot a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k \tag{10}$$

$$c_{Mk}^{SSM} = \theta_{Mk} \cdot c_{Mk}^{binary} \tag{10'}$$

Where coefficients of $\boldsymbol{\Theta}$ are determined as follows:

$$\theta_{Mk} = \begin{cases} 0 \ if \ any \ a_{mk}^{lca} = 0^* \\ 1 \ otherwise \end{cases}, \qquad \begin{cases} \forall \ M, k \\ \forall \ m \in M \end{cases} \tag{11}$$

If there is any a_{mk}^{lca} known to be zero by the practitioner ($= 0^*$), then θ_{Mk} is set to zero. Otherwise the complement is left untouched.

Applied to the production of the pump, if the practitioner is aware that it does not require plastics (m), θ_{Mk} is set to zero with M being the plastic product group and k being the production of the pump, which ultimately results in a null complement for Plastics (c_{Mk}^{SSM}) in the upstream cut-offs (figure 5 b).

4.4.2.3 By-Product Correction

The BPC is also regularly used in HLCAs (Acquaye et al., 2011; Strømman, Solli, & Hertwich, 2006; Williams, 2004). Strømman, Peters, and Hertwich (2009) coined the name for it: By-Product Correction. Instead of setting coefficients to zero whenever they are included into both representations, the BPC consists in subtracting the monetary value of the inputs covered by the LCA inventory from the EEIOA complement. Equation (7) is modified subsequently:

$$c_{Mk}^{BPC} = a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k - \sum_{m \in M} (a_{mk}^{lca} \cdot \pi_m) \quad , \{\forall \ M, k \tag{12}$$

$$c_{Mk}^{BPC} = c_{Mk}^{uncorrected} - \sum_{m \in M} (a_{mk}^{lca} \cdot \pi_m) \quad , \{\forall \ M, k \tag{12'}$$

where a_{mk}^{lca} corresponds to the LCA input m for the process k and π_m corresponds to the price of the LCA input m .

Applied to our example, this method results in a -0.39€/unit coefficient for steel, coming from the uncorrected coefficient (0.36€/unit) from which we retrieved the monetary value of the stainless steel represented by LCA ($\pi_m = 0.75$€/kg), see figure 5 c).

4.4.2.4 Tier-wise Purchase Correction

Strømman, Peters, and Hertwich (2009) introduced in the same paper the Tier-wise Purchase Correction (TPC) to avoid potential negative values introduced using BPC. Instead of subtracting the value of an LCA flow ($a_{mk}^{lca} \cdot \pi_m$) from the direct input flows of M ($a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k$) in the EEIOA complement, it spreads the retrieval over other indirect inputs of M over the value chain. In order to identify the flows from which to retrieve, Structural Path Analysis (SPA) is used (Defourny & Thorbecke, 1984). The monetary value of the LCA representation is subtracted from the top-ranking paths identified by the SPA. For the illustrative example, the path CO_2>Steel>Pump>User is the top-rank path and is visible in the matrix (0.36€/unit). Other paths available would be CO_2>Steel>Fabricated metal>Pump>User for example, but such higher order paths cannot be visualized in a coefficient matrix. The monetary value of the double counted input (0.75€/unit) is retrieved as much as possible from the top-ranked path which results in 0, the remainder (0.39€/unit) is then retrieved from the next highest-rank path and so on. Representing the effect of the TPC in a matrix is impossible for the example we chose, since some of the changes introduced by the method occur beyond the scope of what is visible with a matrix representation. For more insights into how the TPC is applied to our example refer to SI- 2.

Method	LCA	EEIOA	HLCA no correction	a) Binary	b) SSM	c) BPC	d) STAM
	Pump (unit)	Machinery (€)	Pump (unit)	Pump (unit)	Pump (unit)	Pump (unit)	Pump (unit)
Pump (unit)	0	-	0	0	0	0	0
Stainless steel (kg)	1	-	1	1	1	1	1
Plastics (kg)	0	-	0	0	0	0	0
Machinery (€)	-	0.07	0.84	0.84	0.84	0.84	0
Steel (€)	-	0.03	0.36	0	0	$0.36-1*\prod_m$	0
Fabricated metal (€)	-	0.07	0.84	0.84	0.84	0.84	0
Plastic (€)	-	0.001	0.012	0.012	0	0.012	0
Diesel (€)	-	0.001	0.012	0.012	0.012	0.012	0.012
Lubricant (€)	-	0.001	0.012	0.012	0.012	0.012	0.012
Other (€)	-	0.4003	4.8036	4.8036	4.8036	4.8036	4.8036
Added value (€)	-	0.4577	5.1324	5.1324	5.1324	5.1324	5.1324
CO_2 (kg)	0.1	0.064	0.868	0.1	0.1	0.768	0.1

Figure 4-5 Different hybrid inventories for different methods to correct for double counting

4.4.3 New method: Similar Technological Attributes Method (STAM)

We introduce a streamlined variant of the SSM with the aim of facilitating the hybridization of extensive datasets: the Similar Technological Attributes Method (STAM). Like the SSM and the binary method, STAM forces to zero the $\boldsymbol{C}^u$ coefficients from EEIOA product groups (M, K) for which LCA inventory (for the production of products m, k) already records a product flow (a_{mk}^{lca} , $with\ m \in M$). Just like SSM, it also introduces additional matrices to preserve null inputs that are deemed to not result from truncation but rather from the specificity of each technology in the LCA process. However, contrary to SSM, it does not rely on the analysis of each individual LCA process that is hybridized. Rather, the correction for double counting is streamlined in terms of a set of heuristics applicable to broad *categories* (i.e., collections of product groups sharing similar input structures). For instance, all processes for the production of Machinery and equipment, Office machinery, Electrical machinery (product groups from Exiobase) may be deemed to have similar physical characteristics and be regrouped in the same category (*Machines*), as they all require similar inputs (in different quantities) such as plastic, steel, electronic components, etc. Then, based on these heuristics, if all inputs from processes that belong to category $\mathcal{M}$ are deemed to be thoroughly covered in all LCA inventories of processes of category $\mathcal{K}$, this is recorded in matrix $\boldsymbol{\Gamma}$ such that EEIOA complements will be forced to zero for all product groups $M \in \mathcal{M}$ in the hybridization of all processes $k \in \mathcal{K}$.

It results in a coefficient $\gamma_{\mathcal{MK}}$, which can be expressed as follows:

$$\gamma_{\mathcal{MK}} = \begin{cases} 0 \; if \; \mathcal{M} \; respects \; heuristics \; applicable \; to \; \mathcal{K} \\ 1 \; otherwise \end{cases}, \; \{\forall \; \mathcal{M}, \mathcal{K} \tag{13}$$

It may then be recorded in $\boldsymbol{\Gamma}$ that, for the hybridization of all processes that belong to *Machines* (e.g., the pump), inputs that belong to the category *Plastics* are assumed to be thoroughly covered in the LCA inventory and require no complement. Based on what could such category-wide judgements be made? We propose three heuristics:

1) Respect of the fundamental balances: As LCA databases get progressively refined, they explicitly strive to respect conservation of mass, carbon, etc. (Weidema, 2011; Weidema et al., 2013). This increases the confidence that certain categories of null inputs are null to reflect the absence of inputs, rather than as a result of truncation and should therefore not be altered. Thus, if an LCA process is mass balanced, we can reasonably assume that all inputs from categories with the potential to contribute to the mass of the product have been inventoried; and the EEIOA complement should not alter the representation of these flows.

For example, all inputs from the category *Plastics* used in the production of machines (from the category *Machines*) are likely to be embedded in the final product and thereby contribute to their mass. Consequently, all EEIOA complements belonging to this *Plastics* category will be forced to zero in the hybridization of all machines. Conversely, plastic requirements are not embedded in final products of the *Grain* category (e.g., for packaging purposes), and may have been omitted by the LCA practitioner without disrupting mass balance; in that case, inputs from the *Plastics* category would be added as complements to all processes of the *Grain* category.

2) Expert knowledge at database level: Not all inputs that have a mass contribute to the mass of the final product, as multiple usages are dissipative or consumptive. For example, the application of fertilizers in the production of grain is not embedded in the grain (heuristic 1 thus does not apply) and may be truncated even in a mass-balanced LCA process. Nevertheless, through dialogue with database providers, it may be evaluated to what extent they are confident that a certain category of inputs has been appropriately covered: confirming, for example, that all processes from the *Grain*

category have thoroughly been inventoried in terms of all inputs from the *Fertilizers* category. The absence of fertilizers may therefore constitute a specificity of the technology (e.g., organic agriculture) rather than a truncation. This second rule follows the same logic as SSM, but it is applied to large categories of products for speed and efficiency. Such understanding may be directly embedded in the **Γ** matrix.

The introduction of these two heuristics refines the previous definition (11) of $\gamma_{\mathcal{MK}}$ as follows:

$$\gamma_{\mathcal{MK}} = \begin{cases} 0 \ if \ \mathcal{M} \ involved \ in \ the \ fundamental \ balance \ of \ \mathcal{K} \\ 0 \ if \ \mathcal{M} \ deemed \ fully \ covered \ in \ inventories \ of \ \mathcal{K} \\ 1 \ otherwise \end{cases}, \{\forall \ \mathcal{M}, \mathcal{K} \tag{14}$$

$\gamma_{\mathcal{MK}}$ being set to 0 if category $\mathcal{M}$ participates to the mass of category $\mathcal{K}$ or if category $\mathcal{M}$ is deemed to have been thoroughly covered in all inventories of category $\mathcal{K}$, 1 otherwise.

3) Equal coverage within categories: As a third heuristic, we may assume that if an input is present in a process-based inventory, inputs with similar functionalities have been investigated as well by the data collector. Indeed, if an input of diesel has been accounted for, it is unlikely that the data collection omitted to account for inputs from the same functional category (i.e., *Liquid Fuels*) such as kerosene or gasoline. Most likely these were left null because the machine runs on diesel and not on any other functionally competing liquid fuel. Another possibility is that the data collector had no idea which of diesel, gasoline or kerosene is used in reality and assigned all energy requirements to diesel. In both cases though, adding kerosene/gasoline inputs from EEIOA would correspond to a double counting incident. Categories whose associated products have similar functionalities ($\mathcal{F}$) are therefore flagged in STAM. Then, the presence of an LCA input belonging to such categories sets other inputs belonging to that category to zero. Mathematically it corresponds to introducing another matrix Φ (different from $\boldsymbol{\Gamma}$ since both matrices rely on different mechanics and do not have the same dimensions) defined such as follows:

$$\varphi_{Mk} = \begin{cases} 0 \ if \displaystyle\sum_{n \in N} a_{nk}^{lca} \neq 0 \\ 1 \ otherwise \end{cases}, \begin{cases} \forall \ M, N \ \in \mathcal{F} \,|\, \mathrm{M} \neq \mathrm{N} \\ \forall \ k, \ \mathcal{F} \end{cases} \tag{15}$$

where φ_{Mk} is equal to 0 if there is a non-null LCA input n belonging to the product group N which belongs to $\mathcal{F}$ while M also belongs to $\mathcal{F}$. Otherwise φ_{Mk} is equal to 1 and thus leaving the upstream cut-off coefficient unaffected. Both $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$ are then introduced in equation (8), and two concordance matrices ($\boldsymbol{G}, \boldsymbol{H}$) enable the passage between categories, product groups and products:

$$c_{Mk}^{STAM} = (g_{M\mathcal{M}} \cdot \gamma_{\mathcal{M}\mathcal{K}} \cdot g_{K\mathcal{K}} \cdot h_{Kk}) \cdot \varphi_{Mk} \cdot \lambda_{Mk} \cdot a_{MK}^{io} \cdot h_{Kk} \cdot \pi_k \tag{16}$$

$$c_{Mk}^{STAM} = (g_{M\mathcal{M}} \cdot \gamma_{\mathcal{M}\mathcal{K}} \cdot g_{K\mathcal{K}} \cdot h_{Kk}) \cdot \varphi_{Mk} \cdot c_{Mk}^{binary} \tag{16'}$$

An example of $\boldsymbol{\Gamma}$ is given in SI-2 for a better understanding. STAM constitutes a formalized and refined version of hybridization workflows previously discussed in (Majeau-Bettez & Maxime, 2015) and (Majeau-Bettez et al., 2017). The application of all methods is detailed in SI-2.

4.5 Analysis

4.5.1 Assumptions

All the introduced methods can be divided into two different families based on a core assumption on truncations in LCA: the methods assuming that LCA covers at least some inputs completely (Binary, SSM and STAM) and the methods that assume that all LCA inputs are potentially underestimated (BPC and TPC) (figure 6). The binary method assumes that LCA is not truncated on inputs that it explicitly contains in its process-based LCI. Applied to the pump, it means that the stainless steel inputs represent the total requirements of stainless steel of the actual technology, thus it is kept intact. That method requires no additional knowledge of the process to be applied (see figure 6).

The SSM additionally assumes that an LCA process is not truncated on some additional inputs, specified by the practitioner. As such, the position of SSM in figure 6 varies depending on the practitioner's knowledge of the process (thick grey arrow). The more the practitioner knows about the process and has first-hand understanding of what is not required by the process, the more null inputs values they will keep untouched. Notice that if the practitioner has no expert knowledge of the process to be hybridized, $\boldsymbol{\Theta}$ becomes a matrix of ones and SSM is reduced to the binary method. Furthermore, SSM can yield the same results as STAM, but for both to become equal it would require a disaggregation of the categories of STAM into product groups (red dotted lines).

STAM assumes that inputs that are embedded in the mass of the product do not suffer from truncations. If an input would be part of the mass of the product (given the physical attributes of its category) but is not explicitly in the process-based LCI (like the plastic of the fictional pump), that is because that input is null and not because no data are available, therefore it should not be completed with EEIOA data. Unlike the binary method, it requires to design categories and determine which categories participate to the mass of which category to then feed that information to the STAM matrices. In addition, STAM also assumes that if an input of a category whose products share the same functionality is included in the inventory, then other equivalent technologies inputs are not missing from the inventory.

BPC and TPC methods belong to the second family, assuming that all LCA inputs suffer from truncation to some degree and have to be complemented by EEIOA. In addition, they consider that the sum of the values of hybrid requirements of the process (k) for a particular input (m) and its product group (M) must stay equal to the sum of EEIOA's coefficients for that input's product group ($a_{mk}^{lca} \cdot \pi_m + c_{Mk}^{BPC} = a_{MK}^{io} \cdot \pi_k$). The sum of hybrid requirements for the steel of our pump, when expressed in monetary units (0.75€/unit + $\boldsymbol{C}^u$ coefficient from the steel product group) must be equal to EEIOA's inputs for the steel (0.03€/€*12€/unit for Machinery) which gives -0.39€/unit for c^{BPC}. These methods do not require additional knowledge, as the prices should already be known by the practitioner. Their position on figure 6 also varies depending on how much of the system is hybridized, as will be discussed later in the article.

4.5.2 Further development: Database Sampling Estimation of Truncations

Methods from the right or left sides of the dotted line (in figure 6) are based on the two previous extreme assumptions, which could be bridged and refined by a better understanding of the typical coverage of LCA databases for the various categories of inputs in the different product groups. In other words, answering the question: "In average, how much are the inputs of product (m) truncated in LCA processes that belong to product group (K)?" could serve as an empirical foundation to build a middle ground between the two previous extreme assumptions. That is what the Database Sampling Estimation of Truncations (DSET) would strive to achieve and why it is situated *on* the dotted line in figure 6. Mathematically, it would correspond to (1) performing a weighted average of the LCA processes that belong to a given product group and (2) subtracting the monetary value of these average input flows from the EEIOA representation of this product group, the remainder providing an estimate of the typical truncation level in LCA databases for that product group. However, the validity of this estimate depends on the complete coverage of the technologies of the product groups within the LCA database. In other words, it rests on the assumption that the LCA database contains a thorough sample of all technologies for each product group, and that representative production volumes are available, such that, in theory, the two representations should be equal if not for truncation issues. However, for now, even for the ecoinvent database (Wernet et al., 2016), one of the most developed LCA database, the necessary quality data and coverage is not yet there, even for very detailed product groups (e.g., the electricity, steel and cement product groups). Yet, it remains a promising approach as LCA database continue to develop. An attempt to apply the DSET to the electricity processes of ecoinventv3.3 is available in the SI-2.

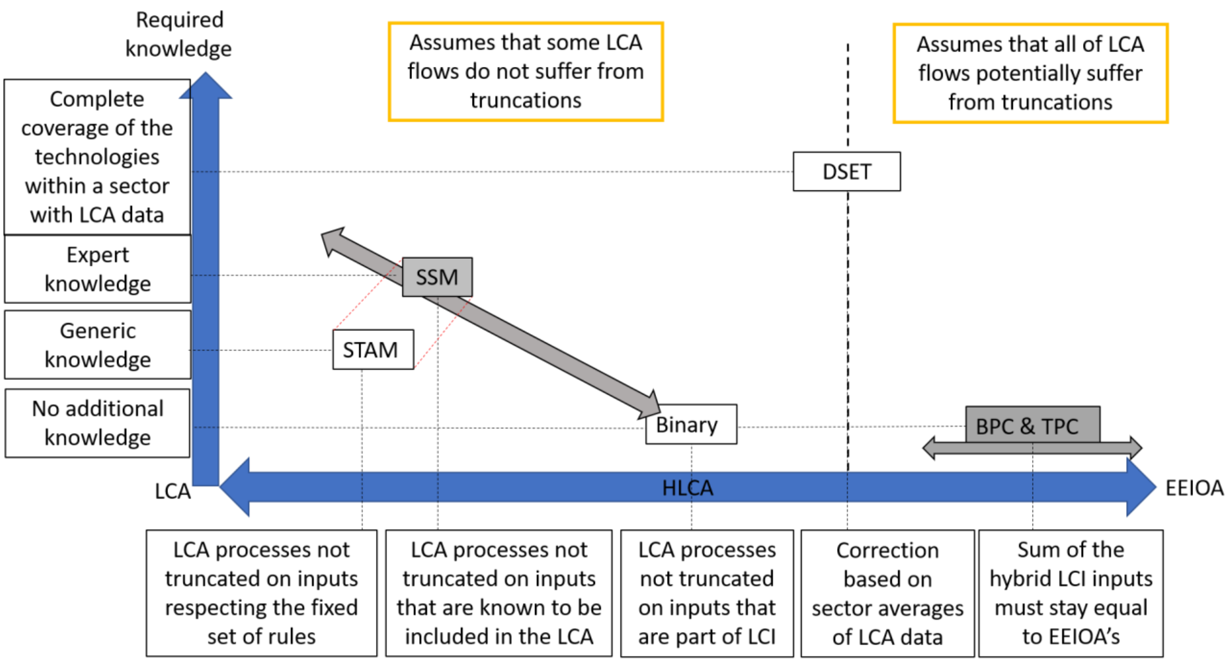


Figure 4-6 Graphical representation to map out relationships between the different methods to correct for double counting

4.5.3 Limitations

Each of the previously presented methods suffers from limitations. The first limitation concerns the incorrect introduction of inputs as EEIOA complements. The binary method relies on the explicit coverage of inputs in the inventory. The problem is that an inventory does not typically contain explicit null inputs. In the example we implemented, it is given that the pump does not require plastics, yet hybridization with the binary method includes plastics in its hybrid inventory (figure 5 a)). The BPC and TPC introduce plastic in the hybrid inventory of the pump too. However, we established that these two methods consider that all inputs of LCA suffer from truncation. It therefore does not correspond to a limitation but to the assumption itself for these two methods. As for SSM and STAM, this limitation is what prompted the implementation of the theta and STAM matrices, which both deal with the issue.

The second limitation of some of the methods is directly linked to how inventories are constructed. In LCA, there is no standard resolution at which to illustrate the product system, and even if the whole supply chain is considered, the different processes can be aggregated depending on the resolution (as illustrated in figure 2), while EEIOA enforces a strict product-group classification. When looking at the hybrid inventory obtained from the binary method (figure 5 a)), the coefficient coming from the fabricated metal product group (corresponding to the prefabricated steel parts of the pump) is left untouched. As a result, part of the steel has been corrected (the coefficient from the steel product group has been set to zero) but part of the steel is still double counted (because the fabricated metal complement is left untouched). The SSM can deal with this issue but it requires the LCA practitioner to be aware of the potential occurrence of this limitation and appropriately modify the theta matrix. In the example we expose, the LCA practitioner is considered not being aware of the issue, which results in the non-correction of the pre-fabricated steel parts too (figure 5 b)). The $\boldsymbol{\Gamma}$ matrix of STAM on the other hand, allows to efficiently correct double counting emerging from the low-resolution of product systems. Indeed, steel and fabricated metal belong to a category (Metals) which is part of the mass of the pump, thus resulting in a coefficient forced to zero for both product groups (figure 5 d)). All the steel is therefore corrected even with minimal understanding. BPC and TPC also deal appropriately with this issue as they subtract the total monetary value of the steel as viewed by LCA (i.e., steel + parts of steel), regardless of the level of aggregation of the inventory.

SSM possesses a practical limitation that other methods do not suffer from: efficiency. To expose this issue, one needs to think about massive hybridization, like the hybridization of ecoinvent v3.3 (about 14,000 processes). It would then require creating a theta matrix for all these processes. More precisely, it requires filling a 200x14,000 matrix manually (using the 200 product groups from Exiobase) and take into account the specific characteristics of each flow, like the problem between steel and steel parts described previously. This issue was the main motivation behind the definition of the STAM.

When applying the BPC, the monetary value of the LCA-covered double counted inputs is subtracted from a single EEIOA complement. If the LCA value of the input is greater than the value of the complement, it will introduce a negative value in the hybrid inventory, like is shown in figure 5 c). This negative value itself does not create any inconsistency for the result of the analysis. However, it has implications when interpreting the results. That is the reason why in the same article, Strømman and colleagues introduced the TPC which was developed to deal with this problem by spreading the retrieval, thus not introducing negative values.

As stated previously in its description, the TPC relies on SPA and subtracts the monetary value of LCA-covered double counted inputs from the top-ranked paths. However, these top-ranked paths can correspond to inputs not in relation with the double counted input. For instance, among the top-ranked paths for the studied pump, there could have been CO2>Steel>Building>Pump>User which corresponds to the steel required by the infrastructure of the location where the pump is produced. In that case, the monetary value of the steel used to manufacture the pump would have been subtracted from another steel: the one used to build the infrastructure. This issue has been highlighted by both the authors (Strømman, 2009) and M. Lenzen (2009).

On a final note, uncertainties are a big problem for all these methods, as well as for hybrid LCA itself. Sensitivity analyses, Monte Carlo analyses and pedigree matrices should ultimately be applied to several aspects of hybrid LCA and of these methods, such as the concordances between products and product groups, the expert knowledge required to apply SSM, the more generic knowledge required by STAM, EEIOA data, price data.

4.5.4 Financial balance in HLCA

Among the introduced methods, three of them force coefficients to zero (binary, SSM and STAM). By doing so, the financial balance of EEIOA is broken, because the sum of a column (A^{lca} inputs, C^{u} inputs and added value) expressed in monetary terms does not necessarily amount to the price of the product anymore. The financial balance has to be re-established, being a requirement of EEIOA and an essential quality check for the validity of an economic process. Yet, when applying these hybridization methods to product flows, practitioners never address this issue. Implicitly it means that the rescaling occurs on the added value (non-product flows) alone, and since added value has no environmental impacts, nothing changes on the HLCA emissions. However, Strømman and Solli (2008) proposed to use the price model of Leontief to rescale the added value and the C^u coefficients instead of only rescaling the added value. Applying this adjustment has an impact on HLCA (figure 7) either decreasing or increasing the EEIOA complements. The price model adjustment considers that an increase in the technology requirements compared to the average technology due to the introduction of specific LCA data in the hybrid inventory should be balanced by a decrease of the EEIOA complements and vice versa. Applied to the binary method for the pump, the latter requires 1kg of stainless steel (0.75€/unit) which is more than is typically required by an average machine (0.36€/unit), thus, applying the price model adjustment means that since more steel is required (from the LCA inventory) and that the price of the pump does not change, fewer materials, services etc. must be used and less profit made. On the other hand, rescaling only onto the added value means that, even if more steel is required, the same amount of materials, services, etc. are used and even less profit is made. The influence of the application of the price model on emissions is shown in the SI-2.

Method	Binary	Binary + Price model	SSM	SSM + Price model	STAM	STAM + Price model
	Pump (unit)	Pump (unit)	Pump (unit)	Pump (unit)	Pump (unit)	Pump (unit)
Pump (unit)	0	0	0	0	0	0
Stainless steel (kg)	1	1	1	1	1	1
Plastics (kg)	0	0	0	0	0	0
Machinery (€)	0.84	0.810	0.84	0.811	0	0
Steel (€)	0	0	0	0	0	0
Fabricated metal (€)	0.84	0.810	0.84	0.811	0	0
Plastic (€)	0.012	0.011	0	0	0	0
Diesel (€)	0.012	0.011	0.012	0.011	0.012	0.013
Lubricant (€)	0.012	0.011	0.012	0.011	0.012	0.013
Other (€)	4.8036	4.6447	4.8036	4.6594	4.8036	5.4393
Added value (€)	4.7304	4.9502	4.7424	4.9553	6.4224	5.7971
CO_2 (kg)	0.1	0.1	0.1	0.1	0.1	0.1

Figure 4-7 Rescaling either of value added (in blue) or of the whole hybridized EEIOA complement (Price model adjustment) to ensure the respect of financial balance in the hybrid inventory (in red).

4.5.5 Hybridization and environmental matrix

There have been recent developments on the number of elementary flows considered in EEIOA (Stadler et al., 2018; Suh, 2009; Yang, Ingwersen, et al., 2017a) leading to a point where some EEIOA tables include more elementary flows than LCA. However, to the best of the authors knowledge, no hybrid LCA has ever hybridized elementary flows and environmental extensions, focusing only on the hybridization of economic flows. Yet, the elementary flows are bound to the technological description and modifying the technological description of a process through hybridization should have consequences on its elementary flows, which should in other words, be hybridized themselves. For instance, if the hybridization of a machine introduces (additional) diesel input flows, this higher consumption of fossil fuels should be mirrored by an increase in direct emissions of the process, notably of particulate matter. Through consistent hybridization, these additional emissions may be estimated based on the emission of particulate matter in EEIOA environmental extensions ($\boldsymbol{B}_{io}$). Conversely, should the environmental extensions not be hybridized, there would be additional consumption of diesel *without* additional emissions of particulate matter, which essentially distorts the production function.

When hybridizing the environmental extensions, the same double-counting correction method used for the technology matrix should be used. The hybridization of elementary flows for the pump results in null complements of CO_2 for the binary method, SSM and STAM, because LCA already covered CO_2. However, for BPC and TPC the EEIOA value of CO_2 emitted is kept, resulting in a 0.768kgCO_2/unit total (figure 5c).

4.5.6 BPC/TPC can cancel hybridization efforts

In most current hybrid LCAs, only some key processes of the system (foreground) are hybridized while the rest of the system is not (Strømman et al., 2009). The BPC and TPC focus on setting the sum of hybrid inputs to equal the sum of EEIOA's inputs. The total value for the inputs of the hybrid LCA will therefore be identical to that of the EEIOA and the same goes for the elementary flows. However, HLCA and EEIOA will still yield different results because the hybrid inventory uses specific LCA inputs instead of the average inputs of EEIOA, meaning that it retains specificity through non-hybridized processes. However, once these LCA processes are all hybridized as well and forced to respect EEIOA proportions, this specificity will be lost. In other words, if BPC or

TPC is applied to a complete process-LCA database and its environmental extensions, calculations relying on the resulting system description will necessarily yield identical results as if they had been performed with the original EEIOA (as demonstrated in the SI-2). That is why in figure 6, the BPC/TPC position can change depending on how much of the system is hybridized, reaching all the way to the extreme right-hand of the scale, that is, a system equivalent to pure EEIOA.

4.6 Conclusion

When choosing between these methods, BPC and TPC should not be selected if the hybridization scope is a full system (like a whole database), but BPC and TPC can be used on a limited number of processes, if their volumes of the flows are uncertain. As for the other methods, it corresponds to a trade-off. The choice between binary, SSM and STAM should be guided by the level of knowledge of the hybridization practitioner about each process as well as the scope of the hybridization. To perform a hybridization without having to obtain additional information, the binary method should be used. If a high-quality hybridization is wanted, SSM should be applied, keeping in mind that it requires extensive time and resources to feed information to the theta matrix. Finally, for a hybridization which requires minimal additional knowledge and yet still delivers more refined results than the binary method, STAM should be applied. It is also important to keep in mind that the binary method, due to its limitations, does not correct all the double counting incidents while STAM corrects double counting incidents more comprehensively. However, since more LCA inputs are left untouched, an HLCA using STAM will be more prone to truncations than an HLCA using the binary method. In other words, STAM takes a conservative approach, and there is a higher likelihood that STAM corrects too much for double counting, whereas there is a higher chance that the binary method does not correct enough for double counting. Since the SSM is a customized approach in which the practitioner has explicit control over which LCA data he keeps or not, with the proper knowledge, the SSM allows to not over- or under-correct.

This paper provided insights about the different ways to deal with double counting incidents in HLCA. We presented four already existing methods (Binary, SSM, BPC and TPC), introduced a new streamlined variant (STAM) as well as a conceptually distinct approach (DSET) that should gain in relevance as the databases cover more and more existing technologies. The article explained how the methods were correcting for double counting with a common notation, as well as their assumptions and limitations through an illustrative example. We also introduced the necessity to hybridize (and thus correct for double counting) the environmental extensions with the same correction method that was used for the technology matrix. Currently the hybrid community only focuses on the hybridization of the technology matrix, leaving the environmental matrix untouched which can trigger inconsistencies (no increase in emissions of particulate matter even though more diesel is consumed).

A recurring aspect of the correction for double counting is the question of null inputs in inventory. Is an input set to zero because it is indeed not used in the technology or is it because no data is available? In the first case, the corresponding complement should be put to zero while it should be left in C^u for the second (for binary-like methods). However, to the best of our knowledge this information is not registered in inventories leaving hybrid practitioners guessing with heuristics, if they are not the ones who gathered the data initially. There is therefore a need notably for database providers to distinguish which data are known to be zero and which data are unavailable.

All methods presented throughout this article can be considered to carry out the correction of double counting (and by extension HLCA itself) in an ad-hoc way. The challenge in HLCA is determining whether the data provided by LCA is truncated or not, and eventually by how much. This could be achieved with a deeper understanding of the uncertainties in LCA, meaning that HLCA could ultimately be treated as a statistical problem.

Finally, we would like to encourage hybrid practitioners to explicitly declare which methods and assumptions were employed in the correction for double counting, whether or not they hybridized the environmental matrix, and how they rebalanced the system (price model or not). HLCA is a complex methodology that requires all of the hybrid community to thoroughly detail studies so that they can potentially be re-used, reproduced or adapted to a different context. A more comprehensive description of HLCA studies will also help practitioners new to hybridization to have a better grasp on the strengths and weaknesses of hybridization leading to a wider adoption within the field, for more complete and detailed sustainability analyses.

4.7 Acknowledgment

The authors would like to acknowledge the financial support of the industrial partners of the International Life Cycle Chair (a research unit of the CIRAIG): ArcelorMittal, Hydro-Québec, LVMH, Michelin, Nestlé, Optel, Solvay, Total and Umicore. The authors declare no conflict of interest and remain solely responsible for the article.

CHAPITRE 5 ARTICLE 2 : HYBRIDIZATION OF COMPLETE PLCA AND MRIO DATABASES FOR A COMPREHENSIVE PRODUCT SYSTEM COVERAGE

Ce chapitre est une reproduction de l'article « Hybridization of complete PLCA and MRIO databases for a comprehensive product system coverage ». Cet article a été publié dans le *Journal of Industrial Ecology* (https://doi.org/10.1111/jiec.12979). Les co-auteurs de cet article sont : Richard Wood, Manuele Margni, Anders Hammer Strømman, Réjean Samson et Guillaume Majeau-Bettez. Dans la suite de cette thèse, cet article sera référencé comme (Agez et al., 2020).

Summary

Process-based life cycle assessments (PLCA) rely on detailed descriptions of extensive value chains and their associated exchanges with the environment, but major data gaps limit the completeness of these system descriptions and lead to truncations in inventories and underestimations of impacts. Hybrid Life Cycle Assessments (HLCA) aim to combine the strength of PLCA and Environmentally Extended Input Output (EEIO) analysis to obtain more specific and complete system descriptions. Currently, however, most HLCAs only remediate truncation of processes that are specific to each case study (foreground processes), and these processes are then linked to (truncated) generic background processes from a non-hybridized PLCA database. A hybrid PLCA-EEIO database is therefore required to completely solve the truncation problems of PLCA and thus obtain a comprehensive product system coverage. This paper describes the construction of such a database using pyLCAIO, a novel framework and open-source software enabling the streamlined hybridization of entire PLCA and EEIO databases. We applied this framework to the PLCA database Ecoinvent3.5 and the multiregional EEIO database EXIOBASE 3. Thanks to the correction for truncation in this new hybrid database, the median and average lifecycle global warming potential 100 of its processes increased by 7% and 14%, respectively. These corrections only reflect the truncations that could be readily identified and estimated in a semi-automated manner; and we anticipate that further database integration should lead to higher levels of correction in the future.

5.1 Introduction

5.1.1 The issue of truncation and challenges in its quantitative evaluation

Process-based Life Cycle Assessment (PLCA) and Environmentally Extended Input Output (EEIO) analyses are complementary methods used to quantify the emissions and impacts on the environment linked directly or indirectly to the use or production of products or services (Jeswani et al., 2010). The two methods possess different strengths and weaknesses. EEIO is mainly used for assessing impacts of consumption (and production) of a basket of goods and is based on national “input-output” inventories recording trading information from companies (R. E. Miller & Blair, 2009b), complemented by environmental extensions developed by various research groups. As a result, EEIO accounts cover the entire economy (Lenzen, 2000). However, the inventories are provided at an aggregated level for national accounting purposes. Information on inputs required by a certain production process is thus complete, but only expressed in terms of aggregated product groups (e.g., motor vehicles). In other words, it is impossible to directly study specific products (e.g., a specific electric car) with EEIO alone (Suh et al., 2004).

PLCA, on the other hand, is designed to support the comparison of specific products (UNEP, 2003). To do so without leaving out any source of environmental impacts would require inventorying all economic and environmental flows through all supply chains relevant for assessing the life cycle impacts. The economy being deeply intertwined (Gibon & Schaubroeck, 2017; Pomponi & Lenzen, 2017), such inventorying is currently hardly achievable, even with databases. As such, PLCA practitioners define explicit system boundaries (International Organization for Standardization, 2006), which allow for a clarification of which processes of the product system are included in the analysis and which are excluded and therefore ignored. These system boundaries trigger two problems: (1) because they are defined subjectively, it is common that two PLCAs with the same focus of study end up not being comparable due to ignoring different parts of the life cycle (Price & Kendall, 2012), (2) the requirements throughout the life cycle left outside the boundaries are ignored even though they have an impact on the environment (Suh et al., 2004). The latter results in either omissions or underestimations of inputs in the unit process. These omissions/underestimations are what is referred to as truncations in PLCA. They constitute a systematic negative bias in estimation of emissions in PLCA studies.

Because PLCA system descriptions are specific but truncated, whereas EEIO analyses rely on complete but aggregated inventories, both methods have been combined in what is known as Hybrid Life Cycle Assessment (HLCA) (Bullard & Penner, 1978). There are several methodological approaches to HLCA (Crawford et al., 2018). In one of the most common approach (i.e., tiered hybrid approach), PLCA requirements are maintained as is, and complemented by the EEIO requirements as additional inputs that are not included in the PLCA (Islam, Ponnambalam, & Lam, 2016).

Multiple studies have contributed to refining the estimation of truncations of PLCA. Junnila (2006) compared results from PLCA with equivalent results from EEIO for five economic services and observed truncations between 50 and 70%. Williams (2004), Zhai et al. (2010), Michelsen, Solli, and Strømman (2008) and Teh, Wiedmann, and Moore (2018) compared results from PLCAs with results from HLCAs and could estimate truncations on particular products (i.e., desktops, photovoltaic systems, forestry and cements) varying from 20 to 75%. Arvesen, Birkeland, and Hertwich (2013) and Wolfram (2015) specifically focused on estimating truncations for renewable energy processes (using HLCA to compare results) and obtained truncations greater than 100%. Ward et al. (2017) estimated that the truncation of services alone could lead to average truncations ranging from 2 to 12% in PLCA processes across multiple economic sectors. Yet, services only constitute a portion of truncated flows, and therefore overall truncation levels are expected to be higher. Recently, Yu and Wiedmann (2018) obtained truncation estimates of 21-32%, after hybridizing the Australian Life Cycle Inventory database with the Australian Industrial Ecology Virtual Laboratory (an Australian Input Output database). Given the wide range of truncation estimates, small sample sizes and geographical limitations found in the literature, *the typical amount of truncation of PLCA is not clearly identified yet. An estimation based on a representative sampling, including processes from multiple economic sectors, across multiple countries and adding complements beyond just services is therefore needed to finally obtain a representative quantification of the typical truncations of PLCA.*

5.1.2 Progress in inventory hybridization and remaining truncation issues

Despite the improvements in completeness achieved by current HLCA studies, in practice most HLCAs still suffer from systematic truncation because of continued reliance on unhybridized process databases. Process databases play a central role in LCA practice: practically all PLCA studies directly inventory processes deemed most specific to the study (i.e., foreground processes) to then model the value chains of these processes via a database regrouping description of generic technologies (background processes). Even if effort is sometimes invested to hybridize foreground processes to minimize the truncations in the estimation of their direct inputs and emissions, only these foreground processes are ever hybridized (Mongelli et al., 2005). A part of the truncations of PLCA is therefore dealt with in HLCAs (for the foreground processes), but some truncation is necessarily left in the unhybridized background.[9] *To reach more complete system descriptions without sacrificing specificity, the LCA community needs a high resolution database of background processes that is itself hybridized with an EEIO database* (see figure 1). Furthermore, as these processes' value chains stretch across multiple geographies, the EEIO database chosen should preferably be a global, MRIO database.

---

[9] In some hybrid approaches, the foreground is hybridized and then linked to an MRIO database, meaning that value chains are modeled relying on IO and not on LCA. For these approaches, the completeness of the value chains is ensured by IO, yet they are modeled through aggregated data. The value chains of such a hybrid study would thus not qualify as being as specific as LCA, unlike what HLCA aims to do, i.e., keeping both completeness *and* specificity.

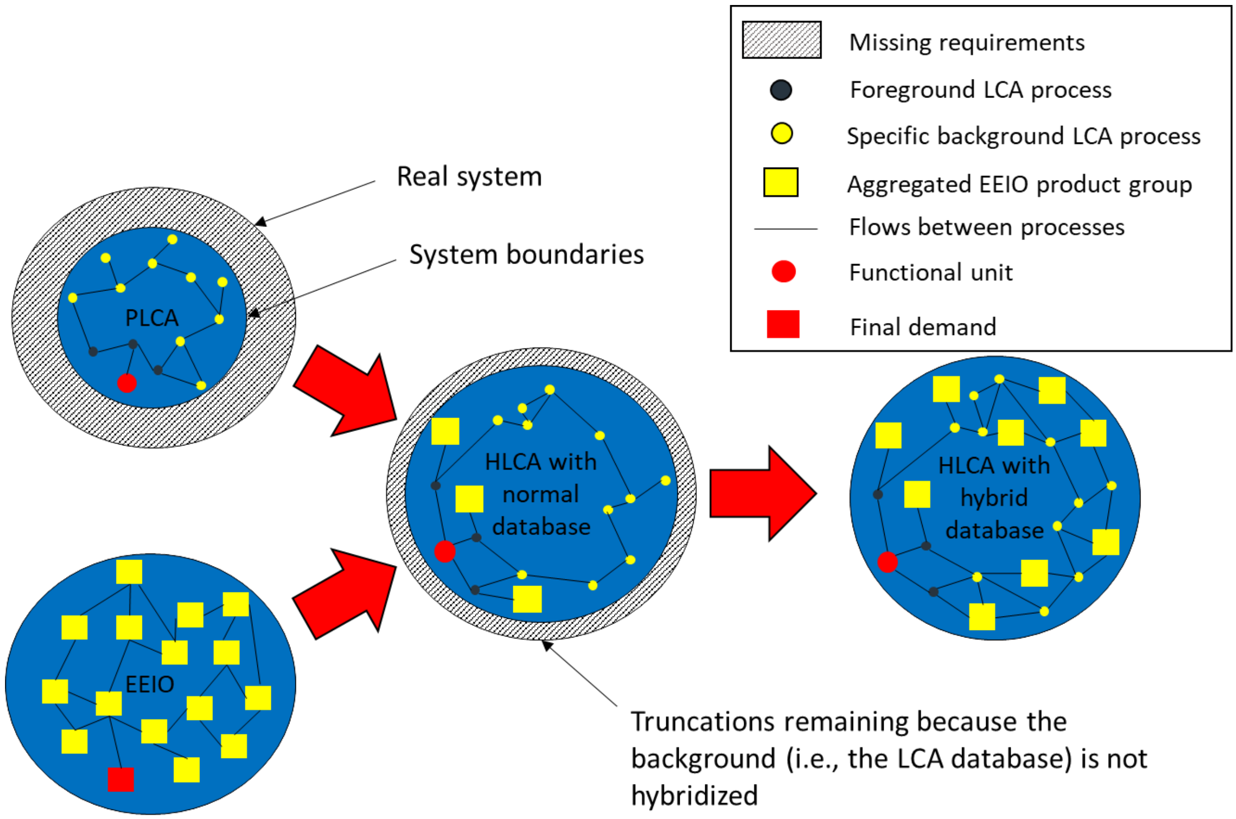


Figure 5-1 Conceptual representation of the hybridization of a PLCA (top-left) and an EEIO (bottom-left) modeling of a product system. Current hybrid LCA only hybridized foreground data (black dots), leaving some truncation in the background. To further remove truncation issues in hybrid LCA, a hybrid database is needed. Only some of the flows between processes are represented for clarity reasons.

5.1.3 Progress and challenges in the development of a hybrid database

The compilation of a hybrid database has long been called for (Bontinck et al., 2017; Crawford et al., 2018; Majeau-Bettez et al., 2011; Strømman et al., 2009)). Previous attempts resulted in one-off hybridizations of a sector-specific PLCA database (Suh & Lippiatt, 2012) and of a country-specific database including multiple economic sectors (Yu & Wiedmann, 2018). To obtain and maintain a multi-sector, multi-country hybridized database, while keeping the specificity of the PLCA database, two broad categories of unresolved challenges must be addressed. First, structural and ontological discrepancies between the economic and process databases (i.e., geography, distribution of goods, intra-establishment flows, etc.) must be addressed preferably in an open framework that is efficient and transparent. Second, the so-called “correction for double counting” must be applied for thousands of processes and should thus preferably be streamlined.

Correcting for double counting corresponds to determining the limits between PLCA and EEIO inputs in HLCA. In other words, which inputs are deemed missing from the unit process inventory (i.e., the processes are fully truncated for these inputs), which inputs are to be kept untouched (i.e., inputs are deemed representative and hence supply chains are not truncated for these inputs), and which inputs should be complemented by EEIO (i.e., the inputs are deemed underestimated resulting in partial truncation). For instance, let us take a detailed but clearly incomplete inventory for the production of a specific car, based on parts listings, public literature, and expert estimates (Hawkins, Singh, Majeau-bettez, & Strømman, 2012). The authors of this study did not have access to required amount of multiple direct inputs (e.g., lubricants, computers or services). Inputs for the production of the average vehicle (from EEIO) could thus be added through hybridization to compensate for this incompleteness. As the EEIO product group represents in an aggregated manner *every* input for the production of the average car, directly adding all EEIO inputs to the PLCA process without appropriate correction would lead to a double counting of the inputs already present in the PLCA inventory. Crucially therefore, such hybridization procedure must always “correct” the hybridization process to avoid any overlap between the requirements inventoried in the original PLCA and the additional requirements estimated through hybridization with EEIO data. For the specific car example, if the authors are confident that their process-based inventory already accounts for all inputs of steel, care must be taken that “steel” and any steel containing part be absent from the additional inputs estimated based on the purchases of the “motor vehicle” product group of an EEIO database, keeping solely the inputs deemed erroneously omitted by the original PLCA inventory (e.g., requirements of business services, research & development, etc.). This careful avoidance of double counting of steel and its associated emissions to the production of the car is commonly referred to as “correction for double counting”.

The current recommendations (Arvesen, Nes, Hertwich, & Huertas-Hernando, 2014; Lenzen & Crawford, 2009; Strømman et al., 2009) for correcting for double counting rely heavily on expert knowledge of the hybridized processes and the limitations of their inventory compilation. These recommendations are therefore impractical for the hybridization of complete PLCA and EEIO databases, for which such expert knowledge of the specificity of each process is less readily available because the hybrid practitioner was not the one who collected the thousands of PLCA inventories. To address this specific challenge Agez *et al*. (2019) developed the Similar Technological Attribute Method (STAM) which relies on heuristics to preserve the specificity of the PLCA database with minimal information about processes. *Until now, however, this approach has been defended solely on theoretical grounds and has not been validated on full-scale databases*.

5.2 Aim and scope

This paper presents (1) the creation of a framework allowing the streamlined hybridization of multi-sector, multi-country PLCA and EEIO databases; (2) the development of a software to ease this hybridization process; (3) the first assessment of the applicability and validity of STAM; and (4) the estimation, across an entire PLCA database covering a wide range of technologies, of the level of truncation that can be readily corrected in process inventories, even in the absence of a detailed, first-hand experience of their compilation.

The streamlined hybridization framework of this paper is applied to the ecoinvent v3.5 database (Wernet et al., 2016) and EXIOBASE 3 (Stadler et al., 2018). By publishing a software tool under open-source license, we aim that any researcher/practitioner with subscription to these two databases can reproduce the hybrid PLCA-EEIO database of this study or, with different parameters, generate a new variant. The focus of this project is to *streamline* the hybridization of a PLCA database, *not* to hybridize each individual process of a PLCA database *perfectly*.

This article solely focuses on complementing missing requirements (truncation) in the LCA database, corresponding to a tiered hybrid approach. Further integration (i.e., disaggregation of IO product groups) is beyond our scope. The reader is referred to (Crawford et al., 2018) and (Suh & Huppes, 2005) for additional information on different variants on HLCA methodologies (i.e., matrix augmented, integrated, path exchange).

5.3 Methods

5.3.1 Database hybridization framework (pyLCAIO)

To complete this analysis, we developed the pyLCAIO framework and software, with key data and processing steps presented in Figure 2. The framework of pyLCAIO can be split into two parts: (1)

pre-processing routines where the different processes of the PLCA database are prepared for hybridization, and (2) hybridization functions where the previously described hybridization process occurs.

The framework of pyLCAIO is solely applicable to PLCA unit processes; so-called “system processes” cannot be hybridized, as they aggregate all the upstream processes whilst the EEIO structure is based on individual processes. The framework requires unique identifiers for each unit process as well as prices for each product, both of which are now generally provided by the PLCA databases themselves, e.g., ecoinvent. The MRIO database must be in a commodity-by-commodity technology matrix format. PyLCAIO requires additional parameters which will be described in the following sections and which are generated manually (in yellow in figure 2).

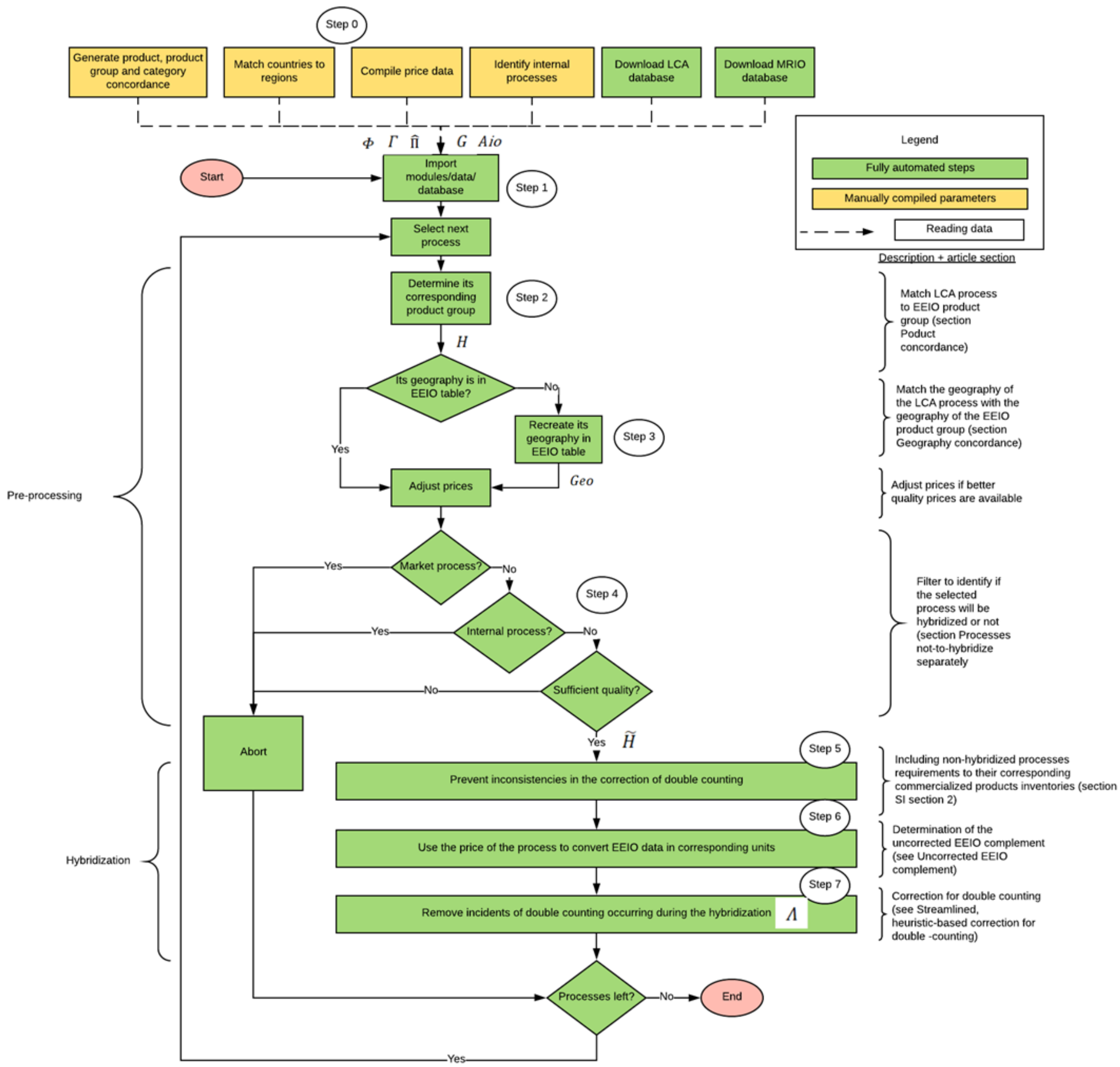


Figure 5-2 Flowchart of the pyLCAIO framework, describing the steps taken (boxes) and intermediate variables (italic letters) that enable the hybridization of PLCA and MRIO databases.$\boldsymbol{\Phi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are matrices required for the correction of double counting. **Aio** is the commodity-commodity technology matrix of the MRIO database. **H** and **G** are concordance matrices, linking products, product groups and STAM categories. $\widetilde{\boldsymbol{H}}$ is the filtered concordance matrix in which processes not-to-be-hybridized are excluded. **Geo** is a matrix dealing with the difference in geographical resolutions of the databases. $\widehat{\boldsymbol{\Pi}}$ is the diagonalized vector of prices. Most steps are automated (in green), although some parameters must be generated manually (in yellow). Necessary data to be provided for pyLCAIO are unit process inventories with

commodity prices for each PLCA products and production volumes for each product group of a commodity-commodity MRIO table.

5.3.2 Pre-processing

5.3.2.1 Product concordance

As a first step, each process of the PLCA database is matched to a single product group of the EEIO database. This matching can be built semi-automatically by relying on standard product and industry code classifications (e.g., CPC, ISIC) and associated concordance tables. The practitioner can then refine this matching based on expert knowledge of the nature of specific processes (Yu & Wiedmann, 2018), in a less automated fashion, i.e., manually (step 0 in figure 2). The matching is then read in by pyLCAIO (step 1) and is transformed in a concordance matrix (**H**, in step 2). If this matrix is populated with ones and zeros to indicate the presence or the absence of a match, each LCA process can only match with one EEIO product group. Cases where the resolution of the MRIO database is greater than the LCA database would require matching each process with multiple sectors in a weighted manner, using coefficients that sum up to one. We did not encounter this situation, which would have to be handled by the practitioner himself based on production volumes and expert knowledge.

5.3.2.2 Geography concordance

When hybridizing, the geographies of the EEIO product groups and the PLCA processes must match as well. For HLCAs of specific case studies, where each foreground process typically has a clear national or subnational geography, it is straightforward to match these geographies to MRIO national classifications. In the development of generic PLCA databases however, processes are sometimes generalized to be applicable in broader geographies (e.g., process representing technologies for Europe or the entire world). Some databases (e.g., ecoinvent) even define dynamic residual geographies all labelled “Rest of the World” (RoW), whose values in terms of countries change for each process. For example, if the database includes two inventories for the production of a given product in Switzerland and in France, then the RoW inventory for this product is considered to represent the average production in all but these two countries.

These different existing regions of the PLCA database must therefore be translated in terms of countries they represent. An initial matching linking basic regions (e.g., Europe) to their corresponding countries (e.g., FR, DE, etc.) must be generated by the practitioner (step 0). PyLCAIO then uses this information to automate two important tasks. First, it identifies which EEIO countries are implicitly represented in the residual RoW geography of each PLCA commodities. For example, the production of lemon is detailed for Mexico, Turkey and Spain in ecoinvent, and therefore "production of lemon, RoW" implicitly represents all regions except these three regions. Second, the framework aggregates national inputs of the MRIO database into the different broader regions of PLCA processes (including the newly defined RoW regions) using the national production volumes of the MRIO database. For instance, an RER process (Europe) will be hybridized with a weighted average of the MRIO database input structure of all the countries in Europe. PyLCAIO finally designs a matrix ($\boldsymbol{Geo}$) matching each process of the PLCA database to the corresponding national product groups (i.e., matching PLCA's "car production Europe" to EEIO's "FR, motor vehicles", "DE, motor vehicles", etc.) with respects to the volume production of each country (step 3).

For cases where the geography of the PLCA process is sub-national (e.g., federal level), pyLCAIO uses the MRIO inputs for the country regardless of the smaller geography described by PLCA. For example, two different technologies from two provinces would be hybridized with the same national average. To be clear, the EEIO complements for these two technologies *might* be identical (depending on the nature of the inputs in their process-based inventory), but their PLCA description will still differ.

5.3.2.3 Processes not-to-hybridize separately

There are types of processes in a PLCA database that should not be hybridized *separately*. These processes should be taken into account during the hybridization of other processes (they are not ignored, see figure 3), but their own direct input structure should not be complemented by EEIO flows.

First, *market processes* should not be directly hybridized. PLCA databases are often structured using two types of processes: production processes and market processes. Production processes describe the transformation of products into other products, while market processes describe the distribution of products into the system (Pauliuk, Majeau-bettez, Muller, & Hertwich, 2016). A

PLCA market process thus (1) aggregates products from multiple providers in a single entity (e.g., “French electricity mix” process aggregating electricity from nuclear, coal, etc.), and (2) adds requirements for the delivery from the producer to the consumer, especially transport. EEIO databases are built in a different way. Market aggregation is pre-calculated by the so-called “construct” algorithms to develop the technology matrix (Majeau-Bettez et al., 2014), and distribution requirements are handled with the “transport margins” and the different price valuations (European Commission, 2008). There are therefore no explicit industry descriptions equivalent to the market processes of a PLCA database in MRIO technology matrices. Hybridizing a market process with a production industry would prove most inconsistent, essentially leading to each process being hybridized twice, doubling the emissions added through hybridization.

Another type of processes not-to-hybridize-separately is what we call *internal processes*. It is common for PLCA databases to divide a production process in multiple sub-processes. These sub-processes are in fact internal to companies, meaning that the companies themselves operate these sub-processes (e.g., a farmer will operate their “tillage” on their own, as one among a multitude of sub-steps to produce their crops rather than hire “tillage services” from a “tillage provider”). In contrast, MRIOs follow the national accounts definition of “establishments” and, as such, only record exchanges across establishments. Consequently, only PLCA processes that lead to final products (exchanged between establishments) are commensurable with EEIO production functions, and only these processes should be hybridized.

We also refrained from hybridizing PLCA processes that are specific to countries not distinctly represented by the MRIO database. Indeed, residual geographies in MRIO databases (e.g., rest of Asia) were deemed too uncertain, as they are generally obtained through extrapolations of other national inventories (Stadler, Steen-Olsen, & Wood, 2014).

Finally, because of the importance of prices in connecting physical PLCA and economic EEIO inventories, there are processes whose level of description or quality is deemed insufficient to support a credible hybridization. These include processes that have no requirements associated to their production (dummy processes); processes whose product has a null or unknown price (e.g., waste treatment); potential instances where the value of outputs is inferior to the cost of inputs; and products whose price exceeds a threshold above which the orders of magnitude of prices are uncertain (e.g., is a mine infrastructure worth 1,000,000€ or 10,000,000€?). In the present analysis, this threshold was set at 100,000€ per functional unit.

Except for internal processes, which must be identified manually (step 0), pyLCAIO automatically identifies all other processes that should not be hybridized (step 4, Fig. 2). This information is used to update the previously defined concordance matrix ($\boldsymbol{H}$), resulting in a matrix $\widetilde{\boldsymbol{H}}$ whose columns are forced to zero for processes identified as not-to-be-hybridized-separately.

5.3.2.4 STAM hybridization categories

As previously noted, most hybridization methods harmonize the boundaries between the PLCA and EEIO inventories through the adjustment or removal *of individual product flows*. In contrast, the core strategy behind the STAM hybridization framework relies on removing or adjusting *broad categories of product flows*. These categories are defined such that the products that they regroup can be expected to respond in a similar manner to a series of inventory quality checks and general heuristics.

Type 1 STAM categories are defined by default so as to encompass product groups whose production share similar input structures (e.g., vegetables, fruits, nuts, cereals, etc. are regrouped in a category "*Agriculture"*), regardless of their varying functionality. Type 2 STAM categories regroup commodities with a similar functionality, regardless of differences in their production functions (e.g., all "*Liquid Fuels"* have a similar functionality). These categories are required for the application of STAM (see section Streamlined, heuristic-based correction for double-counting) and are defined by the user (default categories are available).

5.3.3 Hybridization

5.3.3.1 Uncorrected EEIO complement

Once the pre-processing steps are fulfilled, the hybridization process can begin. First (fig.2, step 6), the uncorrected upstream coefficients ($C^{u}_{\text{uncorrected}}$) are determined following equation (1), which is inspired from (Strømman & Solli, 2008):

$$\boldsymbol{C^{u}_{uncorrected}} = \boldsymbol{A_{io}} \,.\, \left(\widetilde{\boldsymbol{H}} \circ \boldsymbol{Geo}\right) .\, \widehat{\boldsymbol{\Pi}} \qquad (1)$$

where $\boldsymbol{A_{io}}$ is the commodity-commodity technology matrix of the MRIO table, $\widetilde{\boldsymbol{H}}$ is the filtered concordance matrix matching processes of the PLCA database to-be-hybridized to product groups of the MRIO database, $\boldsymbol{Geo}$ handles the disparity in geographical resolution forming EEIO product groups for each region of the PLCA database (e.g., RER, RoW,...), $\widehat{\boldsymbol{\Pi}}$ is the diagonalized matrix of prices for each product of the PLCA database, and $\circ$ represents the Hadamard product.

5.3.3.2 Streamlined, heuristic-based correction for double-counting

To remove the many excessive inputs of $C_{uncorrected}^{u}$ that cause double-counting incidents (step 7), pyLCAIO introduces four filters based on STAM heuristics. These streamlining heuristics are illustrated with a simplified example (figure 3) based on the hybridization of a PLCA process, “chromium steel pipe production” (inspired from Ecoinvent), with the corresponding product group: “Basic iron and steel products” (inspired from EXIOBASE).

During the pre-processing phase, this process was identified as a process to hybridize. It requires inputs from four other processes (among many others) which have been identified as not-to-hybridize-separately, because they are either internal (blue dashed boxes) or market processes (red dashed boxes). Inputs from all these 5 processes are automatically added together in step 5 of pyLCAIO (see SI1 section 2) to be jointly considered during the hybridization (Fig. 3, inputs to green box), through which pyLCAIO strives to avoid overlap between all process requirements (orange) and EEIO inputs (purple flows) in the hybrid system.

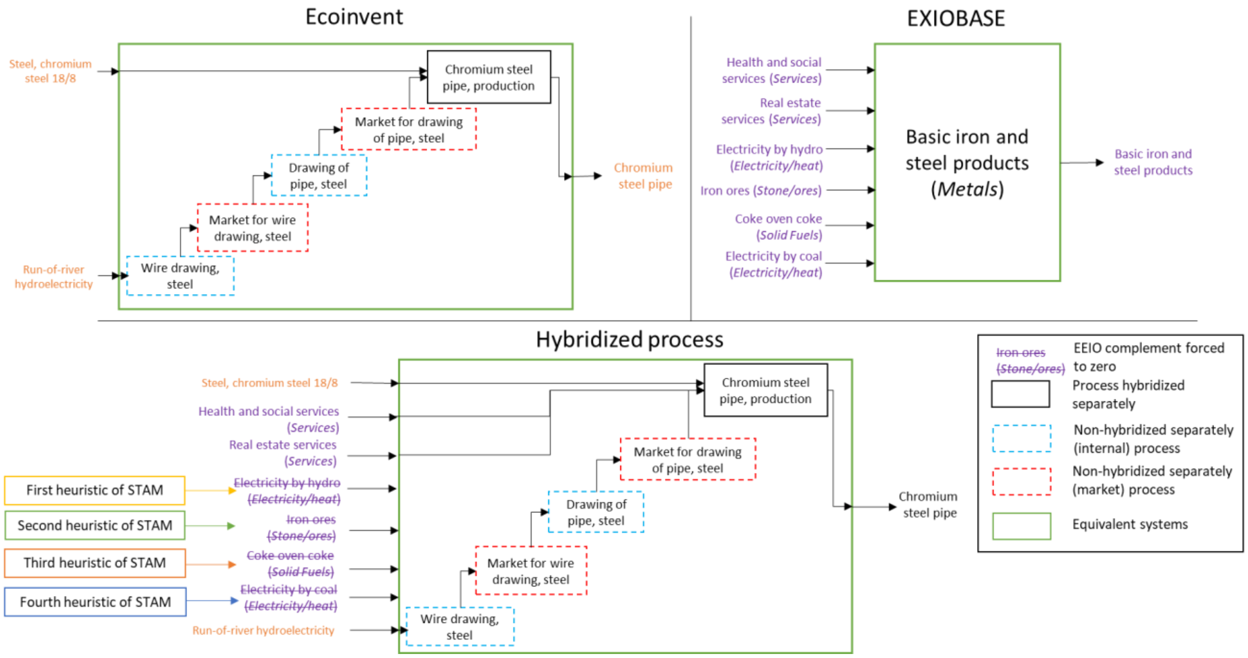


Figure 5-3 Double counting correction illustrated on the hybridization of the 'chromium steel pipe, production' process of ecoinvent belonging to the 'Basic iron and steel products' product group and to the 'Metals' "STAM category" of products. Only a few product groups (purple) and PLCA requirements (orange) are represented here because of space restrictions. Categories to which the different EEIO product groups belong to are specified in the parentheses. Green boundaries represent the equivalence between the PLCA system and the EEIO system. Note that the structure of the ecoinvent process is kept in the resulting hybrid process and that EEIO complements are only added to the process to hybridize separately. The run-of-river hydroelectricity inputs does not appear directly in the unit process inventory of the chromium steel pipe production but is still considered during the correction for double counting (refer to SI1 section 2 to see how).

This first STAM heuristic is based on the simplifying assumption that if the process-based inventory already accounted for one requirement from a given product group, then all requirements from this product group have been fully captured. Consequently, the first STAM filter (Λ) forces to zero all EEIO complements (C_uncorrected^u coefficients) for which a matching PLCA process flow is present. The pyLCAIO software automatically generates this filter matrix of zeros and ones based on product classification concordances (H) and PLCA process inventory data. For example, in figure 3, there is already an input of run-of-river hydro-electricity in the process-based inventory, therefore the EEIO complement “Electricity by hydro” is set to zero.

The second STAM heuristic refines this picture based on the assumption that (unallocated) PLCA inventories strive to respect some fundamental balances, and that an EEIO complement that disrupts these fundamental balances is likely causing double-counting errors. In the present analysis, because of ecoinvent’s emphasis on mass balance in its inventory compilation guidelines (Wernet et al., 2016), the second STAM filter ($\breve{\Gamma}$) removes categories of EEIO inputs that are likely to contribute to the product’s mass and thereby disrupt mass balance. The identification of which inputs likely participate to the mass of which outputs is estimated in terms of broad STAM categories (type 1) by the practitioner in step 0 and recorded in a category-by-category matrix (Γ). In figure 3, since the category Stone/ores typically participates to the mass of the products from the category Metals (to which the pipe belongs to), adding a complement from this category would break the mass balance and “Iron ores” is therefore set to zero.

The third STAM heuristic is based on the assumption that broad product categories can effectively capture the expert evaluation of database compilers concerning trends in inventory coverage. Instead of reasoning at the flow level, we may use STAM categories to record that a broad category of inputs tends to be well covered in the PLCA production inventories of a given category of products, in close dialogue with database compilers. These categories of requirements, if they prove absent from a given inventory, would then be assumed to mark a specificity of this inventory, rather than an omission in need of being complemented by EEIO data. For instance, the use of Solid Fuels as reducing agents in the manufacture of Metals is an aspect to which ecoinvent’s control process is attentive. Thus, because it is ensured by ecoinvent’s control process, an absence of requirements of Solid Fuels in the production of Metals is assumed to represent a particularity of the PLCA process that should then be preserved during hybridization, i.e., not completed by EEIO flows. This third filter ($\breve{\Theta}$) is based on information recorded by the practitioner in a category-by-category matrix (Θ).

Finally, the last heuristic of STAM leads to a fourth filter ($\breve{\Phi}$) to remove from the EEIO complement any input flow that fulfills the same technological function as one of the already inventoried PLCA inputs, even if these EEIO and PLCA input flows belong to different product groups. This last heuristic is based on the assumption that the data provider accounted for all technological inputs of same functionality. For instance, there is already a requirement of Electricity/heat in the unit process inventory in figure 3, which is an input of "run-of-river hydroelectricity". Because "Electricity by coal" has the same function as the run-of-river hydroelectricity, STAM assumes that the absence of "Electricity by coal" is a specificity of the technology; in other words, the absence of coal electricity in the unit process inventory is assumed to not be an erroneous omission, but rather a distinctive real-world characteristic of the process. Categories of products with similar functionalities are thus identified pre-emptively (step 0) in a category-by-category matrix (Φ), following STAM categories of type 2.

Together, these readily applicable four filters extend equation 1 to reasonably ensure that potential excessive EEIO complements are neutralized out of the hybridization process, as per equation 2:

$$\boldsymbol{C^u} = \breve{\boldsymbol{\Phi}} \circ \breve{\boldsymbol{\Theta}} \circ \breve{\boldsymbol{\Gamma}} \circ \boldsymbol{\Lambda} \circ \left[\boldsymbol{A_{io}} \,.\left(\widetilde{\boldsymbol{H}} \circ \boldsymbol{Geo}\right).\widehat{\boldsymbol{\Pi}}\right] \quad (2)$$

5.3.4 Application for transparency and reproducibility

Every step of the use of pyLCAIO to hybridize ecoinvent3.5 (Ecoinvent centre, 2018) with EXIOBASE3 base year 2011 (EXIOBASE Consortium, 2017), is fully documented in a Jupyter Notebook (GitHub link).

In this study, the 200 product groups of EXIOBASE were regrouped in 23 Type 1 and 5 Type 2 STAM categories as detailed in SI2 (folder STAM categories). The data employed by the authors ($\boldsymbol{H}, \boldsymbol{\Gamma}, \boldsymbol{\Theta}, \boldsymbol{\Phi}$) is reported in SI2 (folder STAM data), along with equations through which they are transformed in STAM filters ($\widetilde{\boldsymbol{H}}, \breve{\boldsymbol{\Gamma}}, \breve{\boldsymbol{\Theta}}, \breve{\boldsymbol{\Phi}}$), in SI1 section How to create the filter matrices.

The price data were directly obtained from ecoinvent3.5 as they were the most convenient and reliable available. These prices, however, were compiled pre-allocation whereas pyLCAIO hybridizes the database post-allocation. They might therefore not always be consistent (instances where the costs are above the price), in which case the pre-processing filter ($\widetilde{\boldsymbol{H}}$) automatically excludes these processes from the hybridization. Furthermore, prices of ecoinvent3.5 are based on the year 2005 while EXIOBASE3 is composed of time series ranging from 1995 to 2011. An inflation rate of 1.13 was thus applied to update prices of ecoinvent3.5 to 2011. The products and geography concordance matrices employed by the software are reported in SI2.

The pre-processing filter of pyLCAIO identified 32% of processes in ecoinvent3.5 as "to-hybridize separately", 27% of ecoinvent3.5 processes were identified as market processes, 12% as internal processes, 4% were dummy processes (without inputs or outputs), 14% of processes had a null/unknown price, 1% of processes had a geography not covered separately by EXIOBASE, 5% had a cost of production greater than their price and 5% were above the fixed price threshold (fixed at 100,000€) (see SI2 for details on the filter). STAM was applied by pyLCAIO to hybridize the technology matrices of both databases, constituting a first application of STAM.

In addition to striving to ensure a database-wide consistent hybridization, we strove to address some issues of both databases in a streamlined manner. First, we replaced the global average electricity price provided by ecoinvent with regionalized, basic electricity prices consistent with EXIOBASE monetary accounts. To this end, we divided monetary production volumes of EXIOBASE electricity technologies in each country by the physical production volumes from the UN Energy Statistics Database (resulting prices available in SI2). Second, we strove to remove small inconsistent flows that were introduced in EXIOBASE3 during the balancing steps. These were removed in broad strokes with an EXIOBASE-specific filter, distinct from the more general STAM filters (see SI2).

Moreover, national productions with production volumes of 10M€ and less were judged to be less reliable and have inconsistencies in their EXIOBASE description (e.g., the very small and potentially atypical production of hydroelectricity in Denmark). Ecoinvent processes matching to these small national product groups were thus hybridized not with national production technology but with the more representative average production of the broader region (e.g., ecoinvent's hydroelectricity in Denmark was hybridized with hydroelectricity of Europe). PyLCAIO provides a feature to do this automatically.

5.4 Results

5.4.1 Database-wise truncation corrections

The thousands of hybridized processes in this first version of a hybrid Ecoinvent-EXIOBASE database allows for a representative analysis of the levels of truncation in PLCAs that can be readily corrected in a streamlined manner, i.e., without a detailed understanding of the inventory compilation step of each process. Figure 4 represents the distribution of these corrected truncations as relative increases for four impact categories covered by EXIOBASE extensions: climate change, acidification, eutrophication and human toxicity. The graphs only include hybridized processes, but note that even processes that were not hybridized have their impacts increased through their connections with hybridized processes. The median readily corrected truncation for GWP100 of all hybridized processes is estimated at 7%. A highly skewed distribution leads to an average estimate of 14%, with a standard deviation of 33 percentage points. As the high standard deviation and large difference between the average and the median suggests, estimates per process have a significant variance (from 0% increase to 1,100% increase), although less than 1% of processes display a GWP100 increase above 200%. For acidification, readily corrected truncation estimates are similar to that of GWP100 with a median of 8%, a mean of 17% and a standard deviation of 39 percentage points. Eutrophication has the highest median and average estimates of all four impact categories with respectively 16% and 31% with a standard deviation of 60 percentage points. Finally, toxicity has a median of 14%, a mean of 30% with a standard deviation of 113 percentage points. Data underlying these graphs are available in the SI3.

We found that roughly three-quarters of the readily corrected truncation of the average process was caused by the omission of direct inputs, whereas one-quarter of the underestimation was introduced indirectly by being linked to processes that are themselves truncated. This distinction may be indicative of the relative importance of hybridizing PLCA foreground processes *and* background databases.

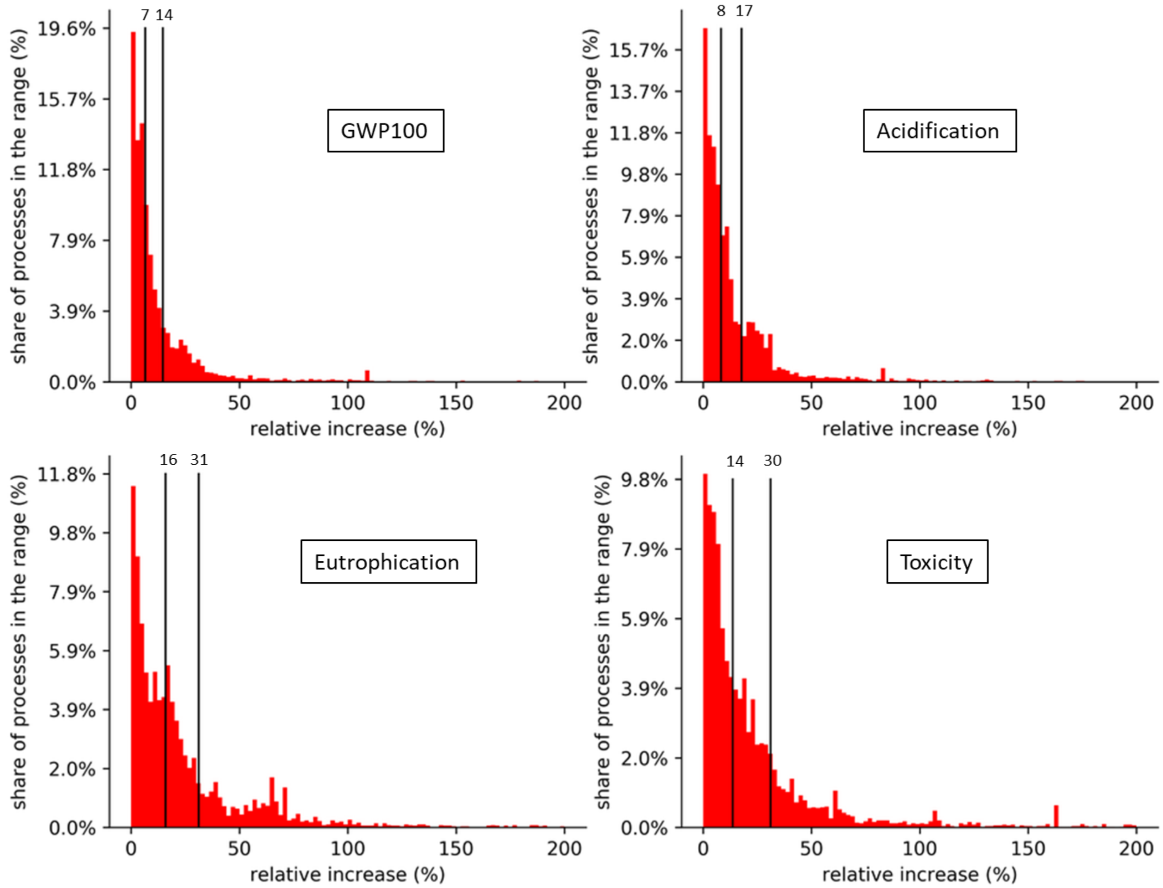


Figure 5-4 Estimations of readily corrected truncation of each process of ecoinvent in relative values (for climate change, acidification, eutrophication and human toxicity) processes of ecoinvent that are hybridized. The medians are 7, 8, 16 and 14% for GWP100, acidification, eutrophication and toxicity respectively. The means are 14, 17, 31 and 30% for GWP100, acidification, eutrophication and toxicity respectively. For each impact category, there are 1 to 3% processes of ecoinvent whose relative increase after hybridization is greater than 200% and are not shown in this figure. Processes that have not been hybridized are not represented in these graphs, even though their impacts increase after hybridization through their connections with hybridized processes. The underlying data behind these graphs can be found in SI3.

5.4.2 Corrected truncation per product groups

Database-wide truncation corrections inform on the general state of truncations in PLCA. Corrected truncation per product groups on the other hand can serve as an indicator as to which group of products need more attention from data providers to cope with truncation issues in PLCA.

All processes from ecoinvent were thus grouped following the product groups of EXIOBASE using arithmetic averages (table 1). We restricted our analysis to product groups for which at least 10 processes of ecoinvent could be identified. Table 1 only shows estimates per product groups for GWP100 (refer to SI2 for the three other impact categories). From table 1 it appears that most of the renewable energy processes are among the most truncated product groups relatively speaking. As these technologies are very reliant on services and capital (Crawford, 2009), which ecoinvent and PLCA as a whole struggle to consider, these high relative estimates were to be expected. Note that even if relatively speaking they are among the most truncated product groups, their absolute GWP increases per kWh are equal to those of fossil energies. Primary resources (e.g., ores, crude petroleum, logs) could be expected to be less truncated than manufactured products, since manufactured products have longer value chains (and therefore supposedly more truncation comes along their production). Some of them however (products of forestry, products of clay, stone), are still estimated to be highly truncated compared to manufactured products, in relative terms. Contribution analyses of a few hybridized processes belonging to the 5 most truncated product groups identified in table 1 can be found in SI2 to get insights as to which EEIO complements are added.

Table 5-1 Relative median, arithmetic average, first and third quartiles, min and max of readily corrected truncations of ecoinvent processes (for GWP100) grouped as product groups of EXIOBASE. Only product groups for which at least 10 processes of ecoinvent could be classified in were assessed. To know to which product group a process of ecoinvent was assigned, refer to the column product group of the truncation_levels tabs in the SI3. To gain insights as to what EEIO complements are added for highly truncated product groups refer to the SI2.

product group	median (%)	min (%)	1st quartile (%)	3rd quartile (%)	max (%)	mean (%)	sample size
Electricity by solar photovoltaic	26	4	24	32	103	29	285
Electricity by wind	16	3	11	23	248	22	300
Products of forestry, logging and related services	14	2	9	25	125	23	86
Motor vehicles, trailers and semi-trailers	13	8	10	21	101	25	24
Wood and products of wood and cork (except furniture); articles of straw and plaiting materials	12	4	10	20	104	18	76
Bricks, tiles and construction products, in baked clay	12	5	8	25	74	24	10
Stone	11	3	4	14	53	12	20
Radio, television and communication equipment and apparatus	11	2	5	24	63	16	59
Machinery and equipment n.e.c.	10	0	5	14	213	16	210
Electricity by biomass and waste	10	0	6	13	173	13	100
Biogasoline	10	3	7	47	84	24	18
Other non-metallic mineral products	9	1	7	14	61	12	48
Chemical and fertilizer minerals, salt and other mining and quarrying products n.e.c.	9	2	5	39	183	30	30
Electricity by nuclear	9	3	5	18	130	16	72
Sand and clay	9	2	5	11	83	18	11
Electricity by hydro	9	0	1	37	376	32	241
Crude petroleum and services related to crude oil extraction, excluding surveying	9	3	5	26	33	15	10
Electrical machinery and apparatus n.e.c.	9	2	6	14	145	15	129
Natural gas and services related to natural gas extraction, excluding surveying	8	1	3	27	108	20	19
Glass and glass products	8	3	3	22	160	18	30
Other non-ferrous metal products	8	3	5	20	151	19	34
Paper and paper products	8	3	6	11	27	9	57
Vegetables, fruit, nuts	8	0	3	16	68	12	122
Collected and purified water, distribution services of water	8	1	3	23	295	39	36
Construction work	7	3	6	9	26	8	75
Fabricated metal products, except machinery and equipment	7	4	5	11	84	16	47
Distribution services of gaseous fuels through mains	7	1	2	80	95	27	14
Other non-ferrous metal ores and concentrates	7	2	4	37	1099	82	33
Chemicals nec	7	0	5	10	131	10	690
Cement, lime and plaster	6	1	2	12	107	11	113
Aluminium and aluminium products	6	1	4	14	91	12	44
Electricity by Geothermal	6	2	4	8	13	6	25
Rubber and plastic products	6	1	3	8	47	8	24
Copper products	6	3	4	11	18	8	19
Basic iron and steel and of ferro-alloys and first products thereof	6	1	4	8	24	7	25
Railway transportation services	6	2	4	7	23	6	22
Steam and hot water supply services	6	0	2	17	371	18	533
Plastics, basic	5	2	4	6	93	10	55
Dairy products	5	3	3	6	7	5	12
Sale, maintenance, repair of motor vehicles, motor vehicles parts, motorcycles, motor cycles parts	4	1	3	8	37	8	18
Cereal grains nec	4	1	3	6	32	5	63
Food products nec	4	1	2	6	11	5	20
Nuclear fuel	4	2	3	5	12	4	51
Wheat	4	1	4	6	8	4	15
Inert/metal/hazardous waste for treatment: landfill	4	3	4	6	7	5	10
Lead, zinc and tin and products thereof	4	2	3	15	41	11	14
Other land transportation services	4	3	3	4	13	4	81
Crops nec	4	2	2	22	220	30	11
Oil seeds	3	0	1	6	8	4	13
Sugar cane, sugar beet	3	0	2	6	32	6	19
products of Vegetable oils and fats	3	1	1	11	21	7	12
Precious metals	3	1	2	3	16	4	15
Natural Gas Liquids	3	3	3	6	10	4	13
Anthracite	3	1	2	9	13	5	13
Textiles	2	2	2	4	37	5	13
Electricity by gas	1	0	1	2	87	2	475
Electricity by petroleum and other oil derivatives	1	0	1	1	32	1	146
Electricity by coal	0	0	0	1	2	1	159

5.4.3 Most influential truncated product groups

To shed light on the omissions that caused the levels of truncation reported in table 1, table 2 shows product groups whose omission triggered the biggest GWP100 underestimation, in relative terms, throughout the ecoinvent database. These are mostly service-related product groups for which current PLCA has limited leverage over. The absence of travel agency services in ecoinvent is what is triggering the biggest underestimation on average (18.7%) in the 5095 processes of ecoinvent that were hybridized directly. The main embedded contributors underlying "Other business services" (responsible for 15.3% of truncations in average) are the energy (70%) and business trips (20%) they involve. Efforts in the data collection on including everything related to business trips (e.g., travel agency, hotel, etc.) would therefore constitute the most promising lead to improving the completeness of the process-based inventories. Other important contributors to readily corrected truncation include missing inputs of software, furniture, printed matter and textiles.

Table 5-2 The 20 economic sectors whose omissions trigger the biggest truncations (in GWP100) in hybridized processes of ecoinvent. For example, the omission of 'Supporting and auxiliary transport services; travel agency services', in average, causes an underestimation of 18.7% in the GWP100 of all directly hybridized processes of ecoinvent.

Exiobase product groups	**Contribution (%)**
Supporting and auxiliary transport services; travel agency services	**18.7**
Other business services	**15.3**
Wholesale trade and commission trade services, except of motor vehicles and motorcycles	**12.4**
Research and development services	**5.6**
Hotel and restaurant services	**4.9**
Renting services of machinery and equipment without operator and of personal and household goods	**4.5**
Financial intermediation services, except insurance and pension funding services	**4.5**
Computer and related services	**3.3**
Retail trade services, except of motor vehicles and motorcycles; repair services of personal and household goods	**3.3**
Furniture; other manufactured goods n.e.c.	**2.9**
Post and telecommunication services	**2.2**
Real estate services	**2.1**
Insurance and pension funding services, except compulsory social security services	**2.0**
Printed matter and recorded media	**2.0**
Food waste for treatment: landfill	**1.3**
Other services	**1.2**
Textiles	**1.2**
Services auxiliary to financial intermediation	**1.2**
Public administration and defence services; compulsory social security services	**1.2**
Bitumen	**1.1**

5.4.4 PyLCAIO

One of the products of this research is the software developed to hybridize Ecoinvent and EXIOBASE: pyLCAIO, which can be downloaded from Github (https://github.com/MaximeAgez/pylcaio), along with documentation, demonstrations, and unit tests. PyLCAIO uses object-oriented programming with Python and is released under a GNU General Public License. PyLCAIO can thus be downloaded, used, modified and uploaded freely. It leverages ecospold2matrix and pymrio (Pauliuk, Majeau-bettez, Mutel, & Steubing, 2015) to read in Ecoinvent and EXIOBASE respectively, but its hybridization algorithm should be generally applicable to any other similarly structured PLCA and MRIO databases.

PyLCAIO thus constitutes a second available platform to facilitate HLCA, the first being the automated path exchange hybrid (APEH) software (Stephan et al., 2018). The two software tools differ by the hybrid methodologies chosen: APEH is based on the path exchange method, while pyLCAIO relies on a matrix approach fit for tiered, matrix augmented and integrated hybrid analyses. Although pyLCAIO is tailored toward the application of STAM heuristics for streamlined database hybridization, other methods to correct for double counting are implemented, providing flexible hybrid analyses.

Currently, pyLCAIO generates the hybrid database in a matrix format which can be exported in a csv format or a pickle format (i.e., a binary Python data storage format). The hybrid database therefore cannot be directly uploaded to mainstream PLCA software (Simapro, OpenLCA, Brightway2) at the moment. Our hope is that pyLCAIO's open-source development strategy will attract a community of practitioners and researchers for its future development and integration to mainstream software.

5.5 Analysis and discussion

5.5.1 Complexity and validity of streamlining assumptions

While correcting for double counting, a non-null PLCA requirement can either be kept untouched (deemed accurate, not truncated at all) or completed with EEIO inputs (deemed partially truncated). In the STAM framework however, a non-null PLCA requirement is always kept untouched (heuristic 1) meaning that we assume a non-null PLCA requirement to never be underestimated. For example, if a process records an electricity input, we assume that this input accurately captures the totality of the direct electricity requirements of the process. In addition, heuristics 2 to 4 also deem some null requirements of PLCA to not be caused by truncation. As a result, the STAM framework is a conservative method, which would rather err toward leaving some remaining truncation rather than overestimating inputs through uncorrected double-counting. This approach contrasts with other methods that are less comprehensive in their correction of truncation and are thus more likely to double count inputs, leading to potential overestimations of hybrid life cycle emissions, e.g., binary method or by-product correction (Agez et al., 2019).

The second heuristic of STAM (non-disruption of mass balance) required particular attention in its implementation, as not all additional EEIO inputs that have a mass necessarily contribute to the mass of the final product. Many products such as pesticides and fuels are dissipated during their use; they might therefore have been omitted even in mass-balanced PLCA inventories and should be kept in EEIO complements. As PLCA databases progress toward recording more product properties, it may become easier to (1) test the extent to which different processes are actually aligned with STAM's assumption that (unallocated) processes tend to include all inputs that contribute to a product's mass because of the respect for mass balance, and to (2) extend STAM's second heuristic to other balances, such as conservation of energy and the different chemical elements. Finally, working at the category level for this heuristic also sometimes wrongly sets EEIO complements to zero, leading to underestimations. It should be kept in mind, however, that the goal of this framework is not to provide an absolute and precise quantification of what is or is not properly accounted for by the PLCA database, but to quickly reason in terms of large-scale, transparent and modifiable set of heuristics.

STAM categories, product concordances and prices used in this study were defined subjectively and only constitute a first step. We expect them to be refined, with the help of a broader community of researchers, throughout the expansion of pyLCAIO.

5.5.2 Outliers analysis

While most of the processes from Ecoinvent display reasonable levels of truncation corrections, there are also about 1% of the processes whose increases in GWP100 scores are greater than 100% after hybridization with EXIOBASE. Most of these high relative increases however, correspond to minor absolute increases. For instance, the production of 1kg of tap water in Canada has a 1.6E-5kgCO_2eq GWP100 score initially and ends up at 5.8E-5kgCO_2eq after hybridization, thus corresponding to a 268% increase. Figure 5 presents the biggest outliers in terms of relative increases among the hybridized Ecoinvent processes that also display non-negligible absolute increases (i.e., > +0.1 kgCO_2eq/functional unit).

Highest relative increases, non negligeable ones (i.e., > +0.1kgCO_{2eq}/functional unit)				
Ecoinvent process	%	kgCO_{2eq}	Justification	Exiobase product group
barite	+1100	+0.2	The price used is 10 times higher than the cited source	Other non-ferrous ores
mint	+220	+0.3	Roughly estimated price	Crops nec
industrial furnace, natural gas	+213	+27500	Proxy process	Machinery nec
solar collector system	+190	+4527	Roughly estimated price	Machinery nec
solar glass, low iron	+160	+1.6	Proxy process	Glass and glass products
selenium	+151	+3.3	-	Other non-ferrous metals
electric motor	+145	+12.2	Roughly estimated price	Electrical machinery nec
powertrain	+140	+25.4	Roughly estimated price	Electrical machinery nec
hexamethyldisilazane	+130	+7.3	Roughly estimated price + is a byproduct	Chemicals nec
pulpwood, softwood	+118	+14.7	Uses price for bentonite	Products of forestry

Figure 5-5 The 10 processes of ecoinvent with a relative correction truncation greater than 100%, resulting in a non-negligible absolute increase of the GWP100 (>0.1kgCO_{2eq}/functional unit). For each of the processes, their corresponding product group (i.e., with which they were hybridized) is provided as well as a possible explanation as to why their relative increase is this high. For all these processes, their respective EEIO product group appear to not be representative of their reality, mostly coming from the aggregation of EEIO.

Typical outliers in our analysis will mainly result from the unrepresentativeness triggered by the aggregation in EEIO. This aggregation can occur both at the technology level (many different technologies regrouped to describe the production of a product, e.g., an “Electricity” product group in typical MRIO databases) or at the product level (many different products and technologies regrouped in a product group, e.g., “Machinery nec” or other not else classified product groups). In both cases, the description provided by EEIO can fail to represent the process of the LCA database, leading to high relative increases after hybridization. Even when an EEIO product group or technology is rather uniform, the EEIO complement can be unrepresentative of the description of a very specific process if this process differs greatly from the rest of its product group. For instance, in figure 5, the “Glass and glass products” is not a very aggregated product group; yet its description does not fit the production of “solar glass” as this technology is atypical among other glass products.

The high level of truncation identified for these processes can also be due to several additional factors.

The results are highly sensitive to price, as there is a linear relationship between price and EEIO complements. In other words, a 10-fold overestimation of the price lead to EEIO complements being overestimated by 10-fold. Such high truncation levels can thus result from the price data of the product from Ecoinvent simply being over-estimated (e.g., for the barite production process) or being roughly estimated (e.g., for the mint production process). Price data for these products can automatically be replaced in the pyLCAIO framework when the PLCA database provider will update its prices or can be directly modified by the pyLCAIO user.

Companies can sometimes include purchases to feed their employees in their declaration. Doing so artificially increases the emissions linked to the product as the employees would still eat even if the product was not manufactured. PyLCAIO offers the possibility to the user to answer this potential issue as he pleases by setting or not categories of consumables as already covered by the PLCA database in the Θ filter matrix. In this analysis, food products purchased by the companies were not included.

Proxies (processes described using another similar unit process of an LCA database) and nearly empty processes (i.e., processes barely requiring any inputs) will inherently have high truncation levels as they are very different from the average of the product group. Proxy processes should not be hybridized as long as the process they are copying is hybridized itself and the filter of pyLCAIO could thus be updated to identify and not hybridize them. For nearly empty processes however, their hybridization could sometimes be justified and thus there is no immediate way to identify which ones should be hybridized.

For these outliers, the streamlined hybridization proposed in this paper might not be adapted, but it only corresponds to less than 1% of ecoinvent's processes. It should be kept in mind also that the goal of this project is to streamline the hybridization and not provide the absolute hybridization of each individual process. Furthermore, pyLCAIO allows the user to exclude these processes from being separately hybridized.

5.5.3 Conservative truncation levels and implications

This study resulted in a representative estimation of truncations with a mean of 14%, which is lower than the isolated estimates found in the literature (between 20 and 100%). Care must be taken however, when comparing these estimates as there are considerable differences between the methodologies used. First, estimates from this paper correspond to truncation levels *once the whole database has been hybridized*, while in the literature only one process is hybridized and then linked to non-hybridized processes. Truncation estimates from this research are therefore more comprehensive and should on average be larger than estimates without a hybrid database. Then, estimates in the literature were mostly calculated following a different method to correct double counting, which typically has a significant impact on results (Agez et al., 2019). We must stress that the STAM method employed in this study aims to preserve the specificity of HLCA processes and deliberately strive for a conservative correction of any double-counting incident, which may unfortunately lead to a potential underestimation of truncation levels overall. Moreover, EXIOBASE does not endogenize capitals in its original version and emissions of products value chains from EXIOBASE are therefore underestimated themselves; in other words, estimates of truncation obtained with endogenized capitals would be higher than the ones of this study. *Estimates from this article are therefore likely to be underestimating the real truncation of PLCA and better represent an estimation of the truncation that can be readily corrected through a streamlined approach*.

The estimations of readily corrected truncation in PLCA may help identify where to prioritize resources in an analysis. Indeed, while hybridization should ideally always be carried out to remove truncation issues in PLCA, table 1 can help deciding when hybridization could be skipped for time and money issues. For instance, comparing a non-hybridized solar power plant process to a non-hybridized wind farm process is more prone to lead to false conclusions, given the high average relative truncations of both processes and close levels of emissions per kWh. Comparing a non-hybridized coal power plant to a non-hybridized gas power plant however, will probably lead to the same conclusions after hybridization.

5.6 Limitations and future challenges

There are limitations to this first deployment of the pyLCAIO framework. First, the hybridization at the supply and use level is not implemented. Currently, pyLCAIO can only hybridize symmetric technology matrices of PLCA and EEIO and cannot process unallocated versions of PLCA database or supply and use tables of EEIO. Working at the technology matrix level triggers some inconsistencies. As mentioned previously, price data for Ecoinvent for example, are compiled pre-allocation and are therefore not always adapted to post-allocation versions of the database. Furthermore, allocation methods used in PLCA databases are not uniform for each process (sometimes monetary, sometimes carbon content, sometimes expert knowledge), while MRIO databases rely on a single allocation assumption to underpin its construct, thus triggering inconsistencies. The goal of this project being to enable the quick creation of one's own hybrid database, we chose to work at the technology matrix level, as there remains restrictions to accessing all unallocated processes of Ecoinvent.

In order to improve the quality of HLCA, additional work must be done to expand the coverage of environmental stressors in MRIO. For instance, EXIOBASE3 only covers about 2% of the environmental stressors of Ecoinvent3.5. It means that the truncation is therefore corrected for only these 2% environmental stressors from Ecoinvent. In other words, main environmental stressors (e.g., CO_2 in air) will be completed through hybridization, but not marginal environmental stressors (e.g., t-Butylamine in water) because the latter are not currently quantified by EXIOBASE. Levels of truncations of this research for toxicity for instance, might therefore still be underestimated as the environmental extensions of EXIOBASE are still lacking.

PyLCAIO currently constitutes a framework incorporating data from an MRIO database to complement a PLCA database. This framework could be further improved to also allow the use of PLCA database data to disaggregate an MRIO database. Such an integration would require accurate price data as well as production volumes, in order to remove the integrated technology from the EEIO production function. Not all the PLCA database has to be integrated into the MRIO database but disaggregating the functional unit from its EEIO aggregated sector would allow *prospective* scenario analyses. For instance, disaggregating electric cars from "motor vehicles" could coarsely model (based on current technologies) what would be the impacts on the economy and environment of an increase in electric car purchases on a national or global scale.

Finally, to better solve truncation issue in PLCA and not rely on ad-hoc methods to correct double counting, data providers would need to explicitly report zero entries. For now, hybrid practitioners are left guessing if a missing requirement is due to specificity (e.g., no pesticides in an organic farm) or due to truncation (e.g., the inventory does not include the packaging of the pesticides).

5.7 Conclusion

The article presented the challenges and proposed answers to the development of a PLCA-EEIO hybrid database. It resulted in a general estimation of 14% for readily corrected truncations in PLCA based on a representative sample. Individual estimations for each product group were also assessed, which can provide insights on the necessity of hybridization to ensure the robustness of comparative studies. This paper also presented the most significant key missing data which can be used by database compilers to direct efforts in improving the accuracy of the descriptions of our technologies.

While hybrid LCAs are considered more accurate than traditional PLCA by resolving truncation/boundary issues (Gibon et al., 2015), they are rarely applied because of the increased requirements in data (mainly prices) and required efforts and expertise to create the framework of hybridization. Thanks to Ecoinvent3 which now compiles price data for its processes (Wernet et al., 2016) and to pyLCAIO which enables non-experts to efficiently perform hybridization with transparent assumptions, these practical issues are finally being resolved. Credibility is perhaps the main remaining barrier to the full adoption of hybrid LCA. We hope that the efforts to keep the specificity of the PLCA whilst providing transparency to the hybridization helps in this endeavour.

5.8 Acknowledgment

The authors would like to thank Guillaume Bourgault from the ecoinvent center for his valuable inputs on the applicability of our methodology to the Ecoinvent database. The authors would also like to acknowledge the financial support of the industrial partners of the International Chair in Life Cycle Assessment (a research unit of the CIRAIG): Arcelor-Mittal, Hydro-Québec, LVMH, Michelin, Nestlé, Optel, Solvay, Total and Umicore. The authors declare no conflict of interest and remain solely responsible for the article.

CHAPITRE 6 ARTICLE 3 : CORRECTING REMAINING TRUNCATIONS IN HYBRID LIFE CYCLE ASSESSMENT DATABASE COMPILATION

Ce chapitre est une reproduction de l'article « Correcting remaining truncations in hybrid life cycle assessment database compilation ». Cet article a été soumis au *Journal of Industrial* Ecology. Les co-auteurs de cet article sont : Elliot Muller, Laure Patouillard, Carl-Johan Södersten, Anders Arvesen, Manuele Margni, Réjean Samson et Guillaume Majeau-Bettez.

Abstract

Hybrid Life Cycle Assessment (HLCA) strives to combine Life Cycle Assessment (LCA) and Environmentally Extended Input Output (EEIO) to bridge gaps of both methodologies. The recent development of hybrid LCA databases constitutes a major step forward in achieving complete system coverage. Nevertheless, current applications of HLCA still suffer from issues related to completeness of the inventory and data gaps: (1) hybridization without endogenizing the capital inputs of the EEIO database leads to underestimations; (2) the unreliability of price data hinders the application of streamlined HLCA for processes in some sectors; (3) the sparse coverage of pollutants in multiregional EEIO databases limits the application of HLCA to a handful of impact categories. This paper aims at offering a methodology for tackling these issues in a streamlined manner, and visualizing their effects on impact scores across an entire LCA database and multiple impact categories. Data reconciliation algorithms are demonstrated on the LCA database ecoinvent3.5, and the multiregional EEIO database EXIOBASE3. Instead of performing hybridization solely with annual product requirements, this hybridization approach incorporates endogenized capital requirements; demonstrates a novel hybridization methodology to bypass issues of price unavailability; estimates new pollutants to EXIOBASE3 environmental extensions; and thus yields improved inventories characterized in terms of 13 impact categories from the IMPACT World+ methodology. The effect of hybridization on the impact score of each process of ecoinvent3.5 varied from a few percentages to three-fold increases, depending on the impact category and the process studied, displaying in which cases hybridization should be prioritized.

6.1 Introduction

6.1.1 Recent progress in hybrid LCA

There are two main techniques to quantify life cycle elementary flows (i.e., the release of emissions to, and use of resources from, the environment) and their impacts: process-based Life Cycle Assessment (LCA) and Environmentally Extended Input Output Analysis (EEIO). Both have their respective strengths and weaknesses. LCA can model the value chain of specific products but suffers from truncation as it necessarily excludes processes from the system boundaries (Majeau-Bettez, Strømman, & Hertwich, 2011; Nakamura & Kondo, 2002; Perkins & Suh, 2019; Pomponi & Lenzen, 2017; Suh et al., 2004). Therefore, the life cycle inventory (LCI) is likely to be underestimated, potentially leading to erroneous conclusions in a comparative LCA (Lenzen & Treloar, 2003). EEIO does not exclude any inputs from the economy , but it only operates at an aggregated level (by industrial sectors), making it impractical for product-oriented analyses (R. E. Miller & Blair, 2009). Hybrid LCA (HLCA) was designed by combining both techniques, each complementing the weakness of the other (Bullard & Penner, 1978). Process-based LCIs are complemented with EEIO data bringing completeness to LCA while EEIO sectors can be disaggregated using LCA data bringing more specificity to EEIO (Crawford et al., 2018). HLCA should thus theoretically solve the truncation problem of LCA while still allowing the study of specific product value chains. Current HLCAs, however, still fail to completely solve the truncation issue of LCA, as the majority of them rely on truncated LCA databases to model the generic value chains (background processes) that are not specific to the system under study (foreground). This issue has been highlighted in literature which has repeatedly called for the compilation of HLCA databases (Bontinck, Crawford, & Stephan, 2017; Crawford et al., 2018; Majeau-Bettez et al., 2011; Strømman, Peters, & Hertwich, 2009)), i.e., entire LCA databases where each process description is completed with EEIO data to mitigate truncation errors. Such HLCA databases have recently emerged thanks to efforts to streamline and semi-automate the hybridization process (Agez et al., 2020; Yu & Wiedmann, 2018). The streamlining enables the hybridization of thousands of processes while respecting a set of rules to ensure the two datasets are reconciled in a coherent manner. By estimating process requirements that had been ignored in the process-based LCAs, the current existing HLCA databases increased the average Global Warming Potential (GWP100) impact score of cradle-to-gate processes by 14% to 32% compared to the non-hybridized version of the same database. Despite this improved completeness, HLCA databases still suffer from multiple limitations leading to some remaining underestimations of elementary flows (Agez et al., 2020), namely (1) the non-endogenization of capital goods, (2) the inability to hybridize certain value chains due to the absence of reliable price data to convert monetary flows into physical flows, and (3) the limited coverage of environmental extensions in Multi-Regional Input Output (MRIO) tables. This article aims to tackle these three issues, which are detailed hereafter.

6.1.2 The place of capital goods in LCA, IO and hybrid LCA

A capital good is a good used in the production of products and services that outlives this production process (e.g., factories, computers, machines). Specifically, in national accounts where product flows are computed on a yearly basis, capital goods are defined as "produced assets which are used repeatedly or continuously in production for more than one year" (European Commission, 2008). In 2007, capital goods formation was responsible for 24% of the total GHG emissions worldwide (C. J. Södersten, Wood, & Hertwich, 2017). Yet, how to account for these impacts with a life-cycle perspective is not uniform across methodologies.

In LCA, capital inputs are entered as direct inputs in unit process inventories, as a "unit" of capital good divided by the estimated total output that unit of capital good will help produce during its lifetime. That way, each functional unit of a process is allocated the same impacts due to capital formation (producing the capital good, e.g., construction of a building) and consumption (using the capital good, e.g., maintenance). For example, in ecoinvent3.5, a photovoltaic cell production plant is expected to produce 4e7m² of photovoltaic cells over its lifetime, and therefore the production of each m² of photovoltaic cell is ascribed a small fraction (4e-7) of this plant's production impacts. The underlying assumption in LCA is that each functional unit is equally responsible for the initial production of the required capital good.

In IO, capital goods are separated from the manufacture of products/services, i.e., the formation of fixed assets is included in the final demand while the consumption of fixed assets is included in the value added. In other words, the production impacts of capital goods are not allocated to the products/services whose production use these capital goods, unlike in LCA. The formation of capital (and its impacts) is computed separately as its own consumption category. To link capital formation (and its associated impacts) to the sectors using them, capital goods must be *endogenized*.

For hybrid LCA, where LCA and IO are to be mixed, consistency on capital goods accounting is paramount, and yet the issue has been largely ignored by the HLCA literature. In the hybrid database of Agez et al. (2020), potentially missing capital goods in LCA value chains could not be estimated with IO, as the capital goods were not endogenized in the IO database. The endogenization of capital goods in IO is seldom applied due to the lack of data on capital goods usage by industries. Nevertheless, taking advantage of a recent surge in its application at national and global scales (Chen et al., 2018; T. R. Miller et al., 2019; C. J. H. Södersten, Wood, & Hertwich, 2018), we are now able to address the following research questions. *How can endogenization of capital goods be integrated in a streamlined hybridization framework? How do endogenized capital goods affect the results of HLCAs?*

6.1.3 The role and limitations of price in hybrid LCA

LCA process descriptions typically compile product requirements normalized per unit of the process's function, and this "functional unit" is typically in physical units. IO production functions also express product requirements per unit of the product or service supplied, but this functional flow is typically (though not always (Duchin, 1992; Merciai & Schmidt, 2017)) expressed in monetary units. In a HLCA, the IO production function is therefore typically rescaled with price data, such that requirements are normalized proportionately to the same physical unit of functional flow as in the LCA inventory.

This role of price data in HLCA leads to two issues. Firstly, if the price of the functional flow is not available (e.g., waste treatment processes in the ecoinvent3.5 database), the hybridization cannot be performed. Secondly, in the case of an LCA process supplying a product whose price differs significantly from the average of the products of its sector, or in the case of highly aggregated sectors where prices are highly heterogeneous (see aggregation bias of IO (Asger Olsen, 2000; Morimoto, 1970)), the use of price data in the rescaling of IO production functions in HLCA becomes unreliable, leading to outliers (Agez et al., 2020; Yu & Wiedmann, 2018).

Relying on prices of functional flows to link IO and LCA thus has serious limitations and can prevent the hybridization when prices for these flows are unavailable or deemed unreliable. For example, in the hybrid database of (Agez et al., 2020), nearly 1,000 processes of the ecoinvent3.5 database could not be hybridized due to price unavailability/unreliability. We will therefore address the question: *How can monetary and physical inventories be hybridized in a streamlined manner* *without relying on price data of the functional flow? Could such an alternative approach allow for the streamlined hybridization of the totality of an LCA database, even including processes with products with uncertain prices?*

6.1.4 The limited coverage of environmental extensions of Multi Regional Input Output tables

Hybrid LCA consists in adding EEIO inputs to introduce requirements deemed missing in LCA, i.e., complement the *technological* description of LCA. The overarching goal of HLCA is to have a more complete estimation of elementary flows resulting from the production of a good or service. Yet, the amount of types of emissions covered by MRIOs is less than that of LCA (e.g., 35 for EXIOBASE3 while ecoinvent3.5, an LCA database, covers about 700). This is due to the priority national accounts have put on covering a given range of environmental stressors consistently across multiple sectors and to the difference in coverage of elementary flows of national accounts. MRIOs thus only select emissions covered by all countries. As a result, HLCA effectively completes emissions of main pollutants such as CO_2, CH_4, SO_2, etc. but is unable to complete more atypical emissions. For example, despite IO complement increasing the inputs of agricultural products taken into account in the value chain of manufactured goods, this increase led to no modification in the pesticide emissions in (Agez et al., 2020). Due to this limited coverage, impacts assessed with MRIO databases are most likely underestimated, especially for impact categories involving a significant number of pollutants such as human toxicity or ecotoxicity (Arvesen et al., 2014). Therefore, the research question to be addressed: *How can the coverage of environmental extensions of MRIO tables be extended in a streamlined manner, such that these environmental extensions become commensurable with that of LCA databases across all impact categories? What difference would this coverage increase introduce in HLCA database impacts?*

6.2 Aim and scope

This research aims to complete the remaining truncations of hybrid LCA databases identified in the introduction through three corrective measures: (1) the integration of endogenized capital goods in the generation of a hybrid inventory, (2) the hybridization of processes without a requirement for (unavailable or uncertain) price information and (3) an exploratory approach to estimate and integrate additional environmental extensions from a national IO table to an MRIO table. For these three measures, we studied their associated effects on 13 environmental impact categories using IMPACT World+ (Bulle et al., 2019). This research builds upon and extends the database hybridization framework developed in (Agez et al., 2020).

We will briefly introduce the methodology used in (Agez et al., 2020). We will then describe how we introduced each new feature in the existing hybridization framework. The effects of these features will then be displayed and discussed. Throughout the article we will refer to the three measures as (1) capital add-on, (2) NFFBH (Non-Functional-Flow-Based Hybridization, see section3.3) and (3) emission add-on.

The research presented in this article was applied to two complete databases: ecoinvent3.5 and EXIOBASE3 (Ecoinvent centre, 2018; EXIOBASE Consortium, 2017). The three measures introduced in this article were applied database-wide (and not to isolated processes) with a systematic, reproducible and transparent framework. The code to reproduce our work is available on Github (https://github.com/MaximeAgez/pylcaio) under an open-source license. Because our aim was to streamline the hybridization of thousands of processes following a set of generally applicable rules, the resulting hybrid database does not necessarily provide the best possible hybridization for each process.

6.3 Methods

In subsection 3.1, we provide a brief description of the hybrid LCA method previously introduced by (Agez et al., 2020). Then, we present three measures to cover gaps in the previously published method: the capital add-on (subsection 3.2), the NFFBH (subsection 3.3) and the emission add-on (subsection 3.4). Finally, we provide more detailed information about our case study (subsection 3.5).

6.3.1 Brief hybrid LCA methodology

Hybrid LCA, as applied in this article, consists in complementing data deemed missing from unit process inventories of LCA using the data structure from IO to estimate these missing inputs. For example, the production of electric car includes no inputs of R&D services or computers, as the original compilation did not quantify these flows, and yet they may be estimated based on the use of services by the “Motor vehicles” sector in IO. It is a two-step process. First, *all* inputs of the corresponding sector from IO are added to the LCA unit process inventory, using prices to convert the monetary dimension of IO to the physical dimension of LCA. Some inputs are thus necessarily double counted. The second step is therefore to correct for this double counting, making sure that every input is only accounted once, either by LCA or IO (Agez et al., 2019).

The streamlined hybridization of a given LCA process is performed following equation (1):

$$c_{i,j}^{u} = \alpha(i,j,k) \times a_{i,k}^{io} \times \pi_j \quad (1)$$

where $c_{i,j}^{u}$ is the IO complement of sector i added to LCA process j, $\alpha(i,j,k)$ is composed of different filters (of values 0 or 1) correcting for double counting incidents, $a_{i,k}^{io}$ is the IO coefficient from the technology matrix giving the purchase of products from sector i by sector k, with sector k being the sector to which process j belongs to and finally π_j is the price associated to the product of process j. For more information on this equation and all the variables, the interested reader is referred to (Agez et al., 2020). Note that the understanding of everything underlying $\alpha(i,j,k)$ is not required to comprehend the research presented in this article.

6.3.2 Endogenizing capital goods in hybrid LCA

The endogenization of capital goods in IO results in a $\boldsymbol{K}$ matrix with the same dimensions as the technology matrix $\boldsymbol{A}$. The coefficient $k_{i,k}^{io}$ of the $\boldsymbol{K}$ matrix details how much capital good i was required this year, for the production of 1€ of products from sector k. Emissions due to capital formation are thus connected to the production of products in IO by adding $\boldsymbol{A}$ and $\boldsymbol{K}$. The traditional Leontief's equation with environmental extensions is thus modified as follows:

$$\boldsymbol{D} = \boldsymbol{C} \cdot \boldsymbol{S} \cdot \left(\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{K})\right)^{-1} \cdot \boldsymbol{Y} \quad (2)$$

where the $\boldsymbol{D}$ matrix gives all total impacts on the environment, the $\boldsymbol{C}$ matrix contains characterization factors, $\boldsymbol{S}$ contains environmental extensions, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{A}$ is the technology matrix, $\boldsymbol{K}$ is the capital matrix and $\boldsymbol{Y}$ is the final demand.

To integrate the endogenized capital goods inside the hybridization framework, equation (1) is modified as follows:

$$c_{i,j}^{u} = \alpha(i,j,k) \times (a_{i,k}^{io} + k_{i,k}^{io}) \times \pi_j \quad (3)$$

where $k_{i,k}^{io}$ is i-th row and k-th column element of $\boldsymbol{K}$.

In this research, we used the endogenized version of EXIOBASE3 developed by Södersten et al. (2020) , as EXIOBASE was also the MRIO chosen by Agez et al. (2020) to produce their hybrid database. Södersten et al. (2020) applied the flow matrix method (Lenzen & Treloar, 2004) to the consumption of fixed capital, using WORLD-KLEMS (WORLD KLEMS Initiative, 2017), the EU KLEMS database (The Vienna Institute for International Economic Studies, 2017) and national accounts to provide data describing capital goods transactions across products and industries.

Since capital inputs from IO are endogenized, they must not be double counted with those from LCA. To follow with the methodology used in (Agez et al., 2020) we relied on the STAM method (Agez et al., 2019) to correct for double counting. This method does not add IO inputs if a corresponding input is already accounted for by LCA. In other words, when a capital good was already included in the unit process inventory of an LCA process, the corresponding k coefficient for this capital good was not added to the unit process inventory, i.e., $\alpha(i, j, k) = 0$.

6.3.3 Non-functional-flow-based hybridization (NFFBH) to avoid price issues

EEIO analysis relies on so-called Leontief production functions, where the ratios between all the inputs and all the outputs are assumed to remain invariably fixed. For example, the ratio between iron ore, coal and labour inputs in the production of steel is assumed fixed, in a rigid "technological recipe". LCA typically relies on similar fixed technology descriptions.

As previously discussed, in combining IO and LCA production functions, the two must be rescaled to a common flow with price data. Traditionally, the common flow that defines this rescaling has always been the *functional flow*, expressed in physical units. The LCA inputs are expressed per kg of steel output, while the IO requirements are rescaled with price data to express the inputs per kg of ferrous metal supplied, for example.

But if the ratio between *all* flows is fixed within each production function, then *any* flow that is thoroughly accounted for in both LCA and EEIO production functions could theoretically serve as the common flow. In the hybridization of steel production, for example, we could very well rescale the EEIO production function such that its direct use of coal in physical unit equals that of the LCA process description. Alternatively, if direct emissions of CO_2 are well accounted for in the two datasets, the EEIO description of ferrous metal production could technically be rescaled such that its direct CO_2 emissions equal that of the corresponding LCA process. In both cases, the rescaled production functions could then be combined in a consistent hybrid analysis.

When should such non-functional-flow-based hybridization (NFFBH) be used? And if not the functional flow, which flow among all the input and output flows should be selected? Let us keep in mind that the objective of hybridization is to estimate as precisely as possible requirements that are missed in an LCA process. Let us also keep in mind that EEIO production functions are the results of the aggregation of the "recipes" of multiple production activities within each sector; and that the more constant the ratio between two flows across these activities, the more confidently the scale of one flow can be used to estimate the other. In other words, when comparing multiple activities of a sector, the magnitudes of some flows will be well correlated with one another, while other flow pairs will not. Therefore, *ideally*, the combination of an LCA and an EEIO production function should be based on the flow that, once converted to physical units, is best correlated to the EEIO flows that are missed in the LCA description.

Unfortunately, the statistical distributions around EEIO coefficients and prices are almost never available. The practitioner must therefore fall back on heuristic to identify non-functional flows that are likely to be well correlated to all other inputs:

1. The flow must be thoroughly accounted for in both the LCA and EEIO process descriptions. The NFFBH technique essentially assumes that this one flow is *not* truncated. This assumption may hold true for easily documented, environmentally important flows (fuel use, etc.)
2. For the flow to be well correlated to other flows and to serve as a good predictor of missing inputs, it should be of relatively high importance in the cost structure or the emission structure of the sector, technologically relevant, and expected to be displayed in similar proportions by all activities that constitute this sector.

For example, if we estimate that prices of products from the "Construction work" sector are highly heterogeneous (due to widely varying levels of value added across projects), then the supply flow of this sector (once converted to physical units) is most likely not an ideal candidate for estimating missing product inputs. We could turn, for example, to the flow of "Cement, lime and plaster", which, being a staple of construction works with more narrowly defined prices, may be more closely correlated with the rest of the inputs. In other words, because of price heterogeneity, this example assumes that the calculated mass of cement used in production is a better predictor of the use of steel, electricity, services, etc. compared to the calculated physical amount of building.

Missing inputs can therefore be estimated by reapplying the ratio "Cement, lime and plaster"/"missing inputs" using the LCA quantity of "Portland cement". Implicitly, it is assuming that the ratio "Cement, lime plaster"/"services" in IO is the same as the ratio "Portland cement"/"services" would be if LCA covered services.

Mathematically, an estimated input $c_{i,j}^{u}$ would thus be determined relying on equation (4) instead of equation (3):

$$c_{i,j}^{u} = \alpha(i,j,k) \times (a_{i,k}^{io} + k_{i,k}^{io}) \times \frac{a_{m,j}^{lca} \times \pi_m}{a_{n,k}^{io}} \tag{4}$$

where $a_{m,j}^{lca}$ is the quantity of flow m (used as the common flow to rescale other inputs with) required for the production of j according to a selected LCA database, $a_{n,k}^{io}$ is the quantity required of products from sector n $(m \in n)$ for the production of the sector k in which is included process j and π_m is the price of flow m.

Following up on the previous example to make it clearer: the determination of the amount of IO complement "services" (i) in the hybridization of LCA process "passive house construction" (j) with the "Construction work" sector (k) is based on the amount of LCA process "Portland cement" (m) used by the "passive house construction" and the amount of "Cement, lime and plaster" (n) used by the "Construction work" sector. Instead of relying on the price of the "passive house construction" (π_j), the hybridization thus relies on the price of "Portland cement" (π_m) to estimate how much "services" the production of the "passive house construction" required.

The same logic can be applied using elementary flows. For example, the service inputs of a "waste treatment landfill" process can be determined using the amount of direct emissions of CO_2, because in this case, the direct emissions of CO_2 probably have a better correlation with other inputs. The estimation of missing input $c_{i,j}^{u}$ would therefore follow equation (5):

$$c_{i,j}^{u} = \alpha(i,j,k) \times (a_{i,k}^{io} + k_{i,k}^{io}) \times \frac{s_{m,j}^{lca}}{s_{n,k}^{io}} \tag{5}$$

where $s_{m,j}^{lca}$ is the amount of elementary flow m used by process j and $s_{n,k}^{io}$ is the amount of elementary flow n used by sector k.

Note that there is no more dependency to prices in equation (5), as elementary flows are compiled in physical units in both LCA and IO databases. The amount of "services" is estimated based on the ratio of direct emissions of, e.g., "LCA CO_2" (m) on direct emissions of "IO CO_2" (n), assuming again, that the ratio "CO_2/services" would be identical between IO and LCA, if LCA covered services. The complete list of common flow/sector combinations used in this study can be found in SI2.

6.3.4 Estimating environmental extensions from national to global MRIOs

Currently, the coverage of different types of elementary flows across all existing environmentally extended MRIOs is highly incomplete. To fill this gap, we need to estimate missing environmental extensions across all sectors and regions based on other (less geographically exhaustive) data sources. LCA databases were briefly considered as a potential data source, but they do not match the technological coverage of MRIOs. Instead, this article explores the use of environmental extensions of national EEIOs to estimate elementary flows for all regions of an MRIO. In other words, the sectors of selected countries with detailed elementary flow accounts are used as "geographical proxies" for similar sectors in other regions with less exhaustive statistics on exchanges with the environment. A parallel can be drawn with the use of geographical proxies in the pedigree approach of LCA (Ciroth, Muller, Weidema, & Lesage, 2016; S. Muller, Lesage, & Samson, 2016).

The expanded satellite matrix of the MRIO table $\boldsymbol{S'}$ is then linked to an expanded $\boldsymbol{C'}$ matching the added environmental extensions to corresponding characterization factors. Both matrices are then introduced in equation (2) as follows:

$$\boldsymbol{D} = \boldsymbol{C'} \cdot \boldsymbol{S'} \cdot \left(\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{K})\right)^{-1} \cdot \boldsymbol{Y} \tag{6}$$

In this study, as a first demonstration, we completed the environmental extensions from EXIOBASE3 adapting the methodology of (E. Muller, 2019) and using environmental extensions from USEEIO (Yang, Ingwersen, et al., 2017a), a national EEIO for the United States. USEEIO was chosen for its comparable list of stressors to LCA databases (several hundreds of elementary flows), its relative recency (2012) and its open-source license. In choosing USEEIO, we had to deal with two aggregation issues, i.e., at the technological level and at the extensions level.

Concordances between EXIOBASE and USEEIO therefore had to be produced to match both sectors and extensions. Only missing emissions to EXIOBASE (e.g., acrolein released in the air) were added and existing emissions were not modified (e.g., CO_2). In cases where categories of pollutants were present in both databases (e.g., NMVOCs), but additional emissions from the same family were explicitly reported in USEEIO (e.g., formaldehyde), we added these additional emissions to EXIOBASE's extensions. Otherwise, if a category of pollutant is only present in EXIOBASE (e.g., HFC) but that USEEIO describes each pollutant of this family specifically (e.g., HFC-116), specific pollutants were not added as they were considered to be already included into the category of EXIOBASE. Adding them would result in double counting.

6.3.5 Case study

In this article, we completed the ecoinvent3.5 cut-off version (Ecoinvent centre, 2018) with the product x product EXIOBASE3.7 matrix, base year 2011 (EXIOBASE Consortium, 2017). The technology matrix of EXIOBASE was compiled using an industry-technology construct. The capital add-on was applied using the product-by-product capital matrix base year 2011 (C. J. Södersten, 2020) . We used USEEIOv1.1 for the emission add-on (Yang, Ingwersen, Hawkins, Srocka, & Meyer, 2017b). Finally, we used the midpoint v1.28, global default values from IW+ (Patouillard, 2019). For simplicity, we use abbreviated names for the impact category of IW+ in our presentation of results (see SI1 for concordance with actual IW+ nomination and units of impact categories).

Three impact categories of IW+ were excluded from this study. The impact on ionizing radiations could not be studied as neither EXIOBASE nor USEEIO track pollutants of this impact category. The land transformation impact category was excluded as EEIO does not keep track of the original state of the land used. Finally, ozone layer depletion was excluded as results obtained for this impact category were not deemed robust enough (refer to SI1 for more details).

6.4 Results

This result section investigates the effects of implementing the three measures presented in this article on the environmental impact results. First, we explore the effects of expanding the environmental extensions of an MRIO on the global economy impact scores. Second, we hybridize an LCA database with an MRIO database accounting for all three additional measures and investigate their effects when compared to impact scores of the non-completed version of the LCA database. Finally, we present the sources for data, results and codes generated to compute the case studies.

6.4.1 Expanded environmental extensions of EXIOBASE

Originally, EXIOBASE3 included 32 different emissions to the air, 2 emissions to water and 1 emission to soil for a total of 35 emissions. After estimation of additional environmental extensions, EXIOBASE environmental extensions included 832 emissions to the air, 832 emissions to water and 137 emissions to soil, for a total of 1766 added emissions. Other environmental extensions from EXIOBASE were left untouched (i.e., mineral resources, water flows and land use) as they were as complete as extensions from USEEIO.

The characterization of these additional 1766 elementary flows affected the impact scores of 10 impact categories of IW+ across all sectors from EXIOBASE. Figure 1 shows the relative increase (in percent) of the impact of the global economy (in 2011), between the original version of EXIOBASE3 and the expanded version.

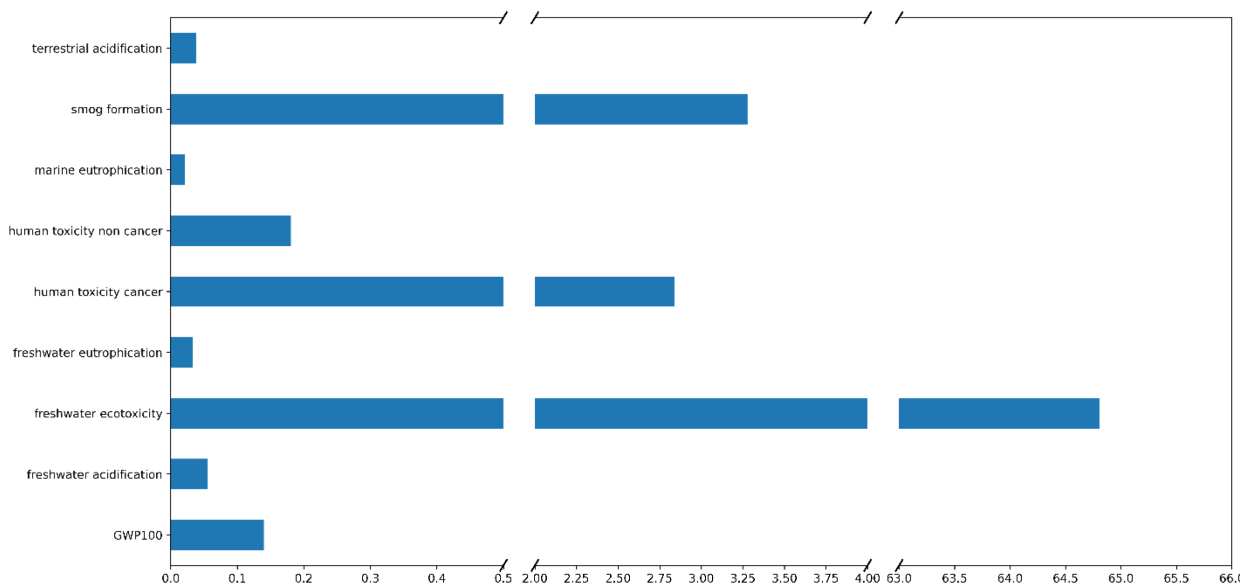


Figure 6-1 Percentage increase for the total impact scores of the world economy (calculated with EXIOBASE3 2011) after estimation of additional environmental extensions based on USEEIO for affected impact categories of IW+. Note that the x-axis was made discontinuous to allow the visualization of all impact categories. Most impact categories are unaffected or marginally

affected by the additional environmental extensions. The total footprint for freshwater ecotoxicity however, is increased by nearly 65%, indicating that EXIOBASE in its original state is not adapted to enable the study of ecotoxicity impacts.

We can distinguish three groups of impact categories. The first group is composed of impact categories for which the increase is lower than 0.2%, namely, terrestrial acidification, marine eutrophication, human toxicity non cancer, freshwater eutrophication, freshwater acidification and global warming potential. Given the limited effect of the additional environmental extensions on these impact categories, we can conclude that EXIOBASE's 35 original emissions, were already covering the most relevant elementary flows. For smog formation and human toxicity cancer, the increase in both impact scores of EXIOBASE is around 3%, mainly due to toluene flows for smog formation and to formaldehyde flows for human toxicity cancer. Finally, the impact score for freshwater ecotoxicity increases by 65%, mainly due to aluminum emissions and various pesticides elementary flows. EXIOBASE in its original state is therefore not adapted to study this impact category. The interested reader can find which emissions were added, classified by impact categories in SI2.

6.4.2 Effect of the measures on LCA impact scores

The corrective measures of this article have different effects. The capital and emission add-ons increase the impact score of each process of the database by a given amount, leading to an overall increase of all impact scores. The application of the NFFBH, in contrast, allows some processes that had previously been excluded in Agez et al. (2020) to now be hybridized. As such, it does not alter the impact score of all processes of ecoinvent, unlike the two other measures.

The effects of the capital and emission add-ons were represented as the relative median impact score increase stemming from the IO inputs added to each unit process inventory, for 13 midpoint impact categories of IW+ (figure 2). In other words, the median of impact score represents the increase when compared to the *original* version of ecoinvent3.5. The blue part of the pie charts represents the increase due to streamlined hybridization relying solely on product flows (following equation 1), as applied in Agez et al. (2020). It puts in perspective the effect of the new measures compared to what was previously achieved. The green part represents the effect of the capital add-on and the red part displays the effect of the emission add-on. Furthermore, we separated the different processes of ecoinvent into 11 categories of products to avoid biases due to the coverage of ecoinvent, i.e., nearly 40% of processes in ecoinvent model electricity generation technologies and markets. The number of processes included in each category is displayed under the name of the category. The application of NFFBH allowed to hybridize 982 more processes. The distribution of these processes is indicated with the nffbh acronym. Note that only 6,082 processes are hybridized, while ecoinvent3.5 includes 16,022, because market and internal processes should not be hybridized to avoid double counting as discussed in Agez et al. (2020).

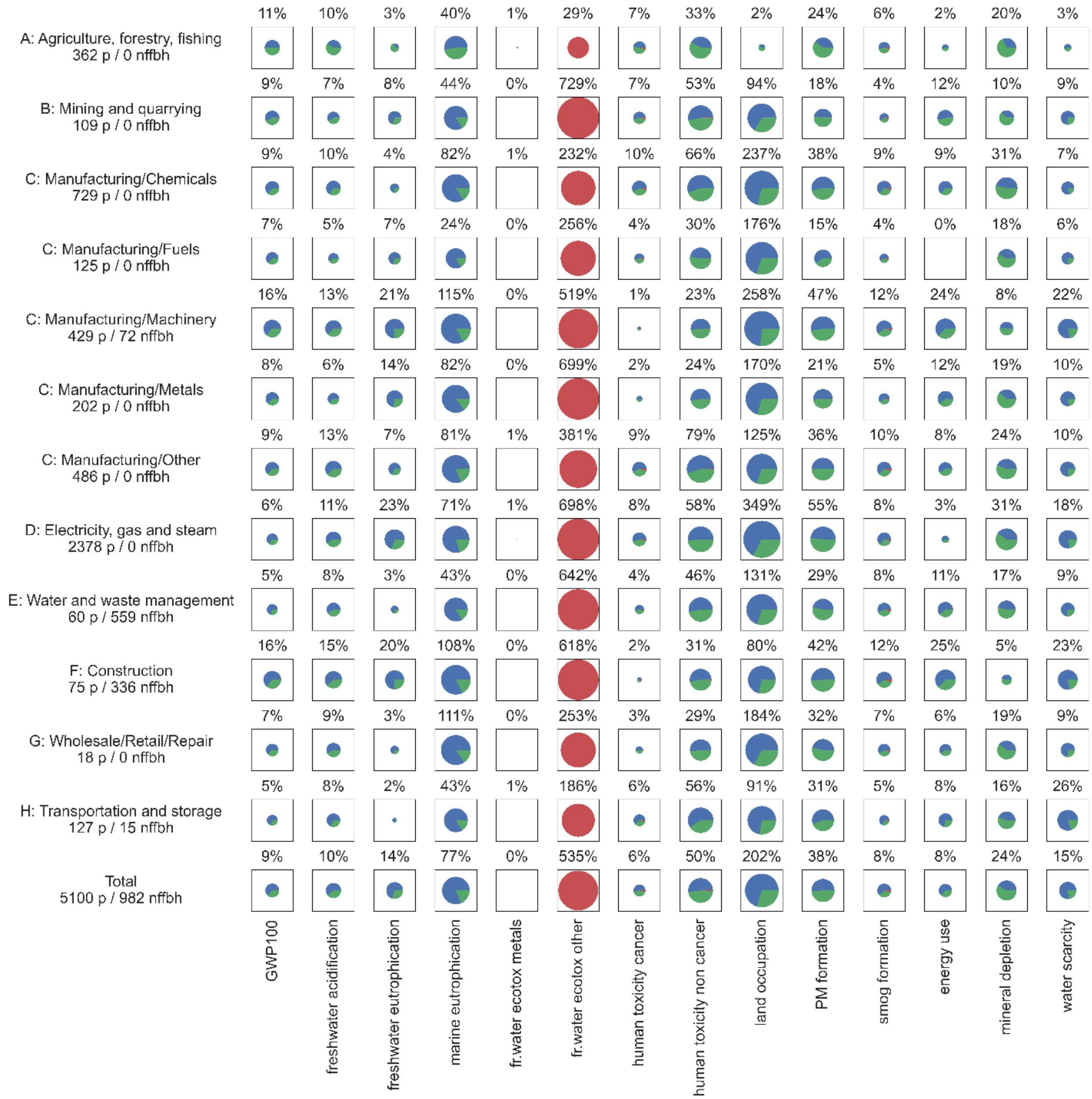


Figure 6-2 Median percentages of impact score increase after hybridization for 13 midpoint categories of IW+ for 11 categories of products. The distribution of the 6082 hybridized processes of ecoinvent across the 11 categories of products is shown under the name of the category. Processes hybridized using the NFFBH are tagged with the nffbh acronym. The median of impact score increase is represented by the size of each pie chart (in logarithmic scale). The blue part of each pie chart represents the increase due to the hybridization as presented in (Agez et al., 2020). The green part represents the increase brought by the capital add-on. Finally, the red part represents the effect of the emission add-on. The application of NFFBH allowed to hybridize

982 more processes, mainly waste treatment processes and buildings which are found in the E and F product categories. Data to reproduce this graph is available in SI3.

Among the 13 midpoint impact categories of IW+ that could be linked to EXIOBASE and its newly estimated extensions, 10 display median increases ranging from 6% to 50% which matches with the expected range of increase due to hybridization found in Agez et al. (2020) and Yu & Wiedmann (2018), though these articles did not cover as many impact categories. We will further explore the remaining 3 categories.

The category which knows the biggest increase due to hybridization is land occupation, with a median total increase of 202%. The disaggregation per category of products shows that agriculture and forestry are only marginally affected for this impact, in contrast with the lifecycle of products from other categories. A major impacting input is therefore missing from value chains of non agriculture/forestry related processes. We know that ecoinvent (and LCA in general) struggles when it comes to introducing services and other seemingly unrelated inputs. One of the services that most companies require and that impacts the land occupation category is the “hotel and restaurant service”. Because of the arable lands that are cultivated to provide the food for this service, the land occupation indicator significantly increases after hybridization.

Marine eutrophication also knows a considerable increase (89%). Major pollutants for this category (nitrogen emissions to water and ammonia emissions to air) come from agriculture and textile activities. Once again, the absence of restauration services in ecoinvent could thus explain this important increase. Similarly, uniforms provided by companies to their employees could also contribute.

The freshwater ecotoxicity category was separated in two sub-categories (one accounting for metal emissions only and the other accounting for all other substances) because metals pose a particular challenge in terms of their coverage in each database as well as the high magnitude and uncertainty of their impacts. Impacts due to metal emissions are only marginally increased by the hybridization while the impacts due to other pollutants (e.g., pesticides, aromatic compounds) are significantly influenced by hybridization.

For LCA practitioners specifically, figure 2 displays the relative importance to perform hybridization with (or without) the measures proposed in this article, and thus should help prioritizing efforts in using hybridization with or without these measures. If the land occupation impact of the production of potatoes is being studied, the hybridization should only have a marginal impact on the results. If the land occupation of the production of coal electricity is studied however, not hybridizing should result in a significant under-estimation of the impact.

Throughout this section we chose to operate with median increases only, as they are more representative than the average increases. That is because the average is heavily influenced by outlier values. In our case study, across each impact category we have several points reaching thousands level of increase which unjustifiably inflate average values. The streamlined hybridization framework presented in this article obviously fails for these processes. However, as was pointed out in the aim and scope section, our goal is not to perfectly hybridize all processes of ecoinvent, rather, our goal is to hybridize thousands of processes in a streamlined fashion. Figure 3 represents the distribution of values for the increase across all ecoinvent processes through violin plots in logarithmic scale. Average values are represented by black crosses, median values by black dots, 1st and 3rd quartiles by black bars. We can see that average values are regularly greater than 3rd quartiles.

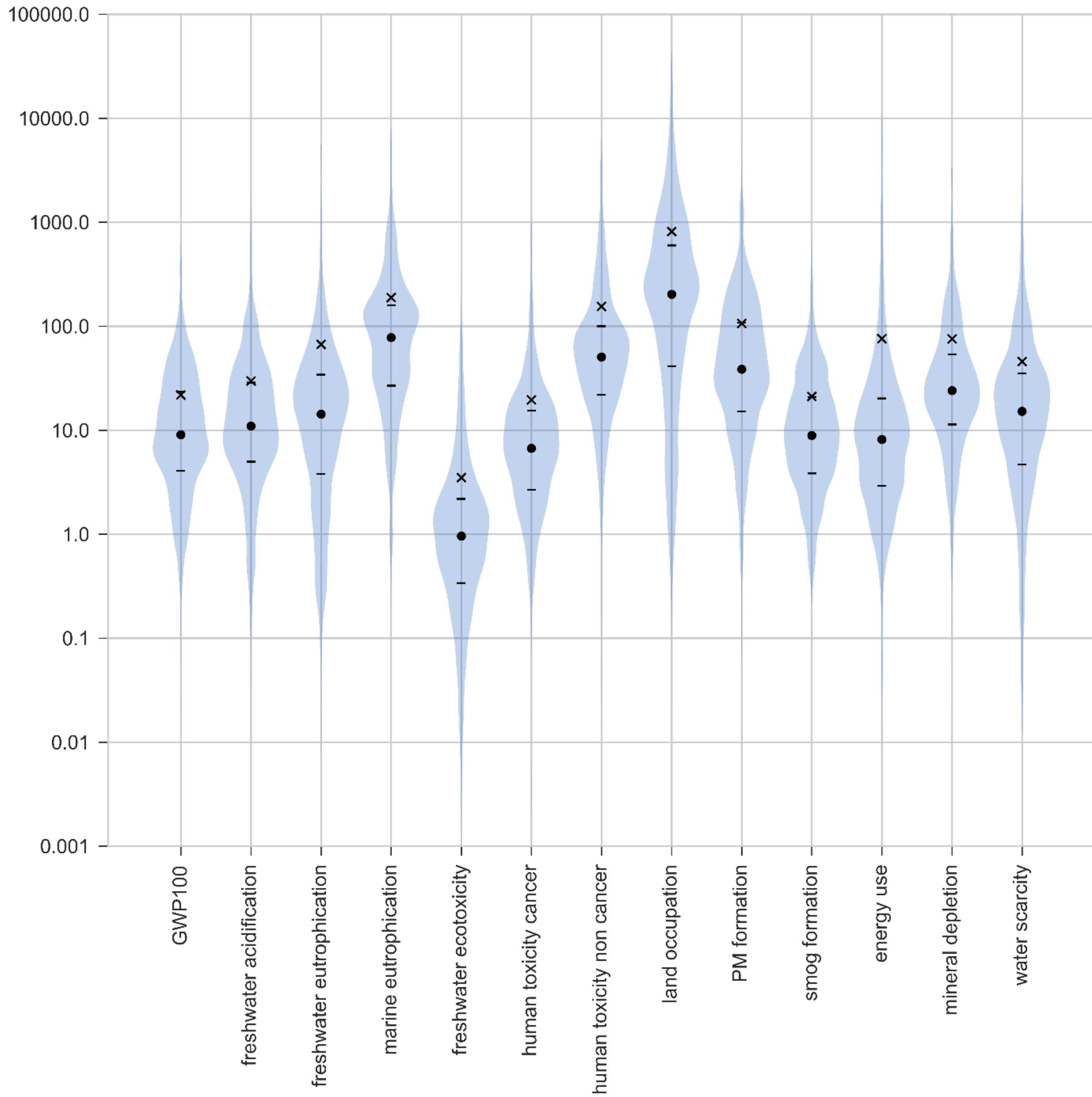


Figure 6-3 Violin plots in logarithmic scale representing the distribution of impact score increase values for all ecoinvent processes. Medians are represented as black dots, means as black crosses, 1st and 3rd quartiles as black bars.

6.4.3 Datasets produced in this study

This study resulted in the creation of four datasets: (1) the upstream cut-off matrix $\boldsymbol{C^u}$ as calculated by equation 6, which contains all estimated missing inputs (e.g., services) for each process of ecoinvent3.5 (https://zenodo.org/record/3890379), (2) a characterization matrix linking the environmental extensions of EXIOBASE3 to the IW+ impact assessment method (https://zenodo.org/record/3890339), (3) the expanded environmental extension matrix of EXIOBASE3 (estimated using USEEIO's environmental extensions) and a characterization matrix linking the added flows to IW+ (https://zenodo.org/record/3890321), (4) a characterization matrix linking ecoinvent3.5 elementary flows to IW+ characterization factors (https://zenodo.org/record/3890367).

The code for completing ecoinvent3.5 with EXIOBASE3 relies on pylcaio v2.0 (https://github.com/MaximeAgez/pylcaio). To use pylcaio v2.0, access to ecoinvent3.5 and EXIOBASE3 databases is required. The software is under a GNL open-source license. It can thus freely be copied, modified and re-used under the same license. The software does not include a user-interface nor does it include features to allow the user to manipulate the resulting hybrid database easily. The use of the database to perform HLCA comparative studies thus requires data manipulation knowledge with the python module pandas (https://pandas.pydata.org) and knowledge of the computational framework of LCA (Heijungs & Suh, 2002). Some of the steps needed to use the resulting hybrid database are described here (https://zenodo.org/record/3890379). In the future, we plan to enable the export of the HLCA database to the brightway2 (https://brightway.dev) software to ease its use for LCA practitioners.

6.5 Discussion

6.5.1 Limitations of estimating environmental extensions from USEEIO to EXIOBASE

In the result section, it was shown that the estimated environmental extensions were paramount to the study of the freshwater ecotoxicity impact indicator while using EXIOBASE. It also appears that a non negligible quantity of smog formation and human toxicity cancer impacts is missed without the estimated environmental extensions. A better coverage of environmental extensions should thus be prioritized by MRIO databases. The methodology presented in this article however, has some overbearing limitations. By re-applying US emission factors to other countries, it is assumed that technologies from other countries have the same performance than in the US. Moreover, regulations also play a disruptive role in this estimation. Some products (e.g., pesticides) used in the US are banned in other countries and vice versa. By simply applying US emission factors to these countries, it is assuming the same products are used in the world.

Some improvements to the methodology could thus be performed. For instance, a reference country per continent, based on a few well-developed national IO tables, could be used to get a better estimation, e.g., estimate Vietnamese emissions based on China's national EEIO. Another lead could be to rely on national emissions databases to identify sectors in countries that tend to pollute more (or less) than US sectors and apply a multiplicative factor to the estimated extensions.

6.5.2 Limitations inherent to the methodology in Agez et al. (2020)

Even though this article strived to correct some of the limitations of Agez et al. (2020) through the addition of the three measures introduced, this research still inherits some of the limitations of the previous article: this work relies heavily on matching products or elementary flows between different databases and is therefore prone to errors; the prices used came from ecoinvent and were compiled pre-allocation while we hybridized the ecoinvent cut-off version, prices might therefore not be adapted; this project does not include any uncertainty analysis.

6.5.3 The case of ecotoxicity

As shown in figure 2, the hybridization does not influence the freshwater ecotoxicity in the same fashion when looking at metal emissions or other emissions. Indeed, inherently, both ecoinvent and EXIOBASE/USEEIO do not put the same focus on these pollutants. We compared the share of metal/other pollutants, scaled up to the whole economy (results in SI2). Obviously, ecoinvent does not cover the whole economy, nevertheless, the distribution of the covered pollutants should stay representative. 99.99% of the ecotoxicity impact of ecoinvent stemmed from metal emissions. The combined EXIOBASE/USEEIO database, on the other hand, calculated that only 33% o f the total impact was caused by metal emissions. Since both impacts were assessed using the same impact method, biases from characterization method should be minimal. There is therefore an inconsistency in the treatment of ecotoxicity between ecoinvent and EXIOBASE/USEEIO. In other words, either ecoinvent covers metals significantly better than other pollutants or EXIOBASE and USEEIO significantly underestimate the amount of metal emissions. The latter can be an inherent product of IO methodology however. Since IO tables are focused on annual data, they do not include long-term emissions and most likely only cover "instantaneous" metal emissions. The coverage of ecotoxic pollutants is therefore an area of improvement for data collection in both LCA and EEIO databases.

6.5.4 Improving capital inputs in LCA

In this study, STAM was used to correct for double counting to keep the methodology previously used in Agez et al. (2020). One of the core assumptions of STAM however, is that if a product is already covered by the LCA database, the data will not be complemented by IO inputs. In other words, if a capital input is already provided by ecoinvent, it was not complemented with EXIOBASE capitals inputs. The endogenized version of EXIOBASE however, provides a more detailed coverage of capital inputs. The LCA community could therefore extrapolate these capital inputs to improve the capital goods coverage in LCA. To enable the extrapolation, an average capital formation and consumption over a period (using EXIOBASE time series for example) would be needed. This could result in better estimates of capital inputs in LCA databases.

6.5.5 Regionalization of impacts in EXIOBASE

This project allowed the linkage of the different datasets used to IW+. This impact method enables the regionalization of impacts. For specific impact category, a pollutant emitted in Australia could impact more (or less) than the same emission in China. One of the struggles of using regionalized characterization factors is that the inventory must be spatialized, i.e., a location must be attributed to the emission that takes place (Patouillard et al., 2018). In LCA, most of the time, the location of the emission is not precise, as many processes are defined for broad regions (e.g., Europe) or as global processes. For MRIOs on the other hand, such spatialization could mostly be done at the national level, as emissions are given for each sector in each country. The use of regionalized characterization factors would thus be interesting in IO to increase the accuracy of impact scores.

6.6 Conclusion

This article completed on the previously developed method of Agez et al. (2020) by introducing three corrective measures: the endogenization of capital goods and their integration into the hybridization framework, the hybridization of processes with unreliable/unavailable prices and the estimations of additional environmental extensions. These three measures further increased the estimated truncation level of LCA previously found in Agez et al. (2020). This study looked at 13 impact categories, using the IMPACT World+ impact method, instead of the 4 categories used previously. It allowed us to see that hybridization is paramount when looking at particular impact categories, e.g., the land occupation impact category saw a 202% median increase across all ecoinvent processes. This study thus helped practitioners determine when hybridization should be prioritized and if including the measures of this article is relevant. Even if the method used in this study to estimate new environmental extensions for EXIOBASE is limited, the results still showed that EXIOBASE is currently not equipped to study the freshwater ecotoxicity impact categories.

The LCA and IO communities have mostly walked their own path when it comes to system description and database compilation. Hybrid LCA and this article try to bring these two communities together in using the strength and shortening the weakness of one another. Potential leads identified during this research are the improvement of capital goods representation in LCA based on capital goods representation in IO, the implementation in IO of regionalization characterization factors used in LCA and the need for both communities to enhance the coverage and characterization of ecotoxicity pollutants. LCA and MRIO database compilation teams would thus greatly benefit from working together.

6.7 Acknowledgment

The authors would like to thank Christine Hung for her assistance on the production of the graphs of this article. The authors would also like to acknowledge the financial support of the industrial partners of the International Chair in Life Cycle Assessment (a research unit of the CIRAIG): Arcelor-Mittal, Hydro-Québec, LVMH, Michelin, Nestlé, Optel, Solvay, Total and Umicore. The authors declare no conflict of interest and remain solely responsible for the article.

CHAPITRE 7 AUTRES CONTRIBUTIONS DE CETTE RECHERCHE

7.1 Développement d'outils

Lors de la réalisation de cette thèse, j'ai contribué au développement de plusieurs outils.

7.1.1 PyLCAIO

Le premier outil auquel j'ai contribué est pyLCAIO, le logiciel permettant l'hybridation de bases de données ACV et IO dont la méthodologie et les attributs sont décrits dans le deuxième article de cette thèse (chapitre 5).

7.1.2 Ecospold2matrix

J'ai également contribué de façon marginale au développement de ecospold2matrix. Il s'agit d'un module Python transformant des milliers de fichiers ecospold (venant d'ecoinvent) en une matrice technologique utilisable dans Python. Ce module est utilisé tout au long de cette thèse pour pouvoir utiliser ecoinvent dans Python.

7.1.3 Brand Emissions Estimator

Le dernier outil développé pendant cette thèse est le « Brand Emissions Estimator » (BEE) de Climate Neutral (https://bee.climateneutral.org/), pour lequel j'ai développé le back-end du logiciel, i.e., l'aspect calculatoire du logiciel. Le rôle du BEE est de permettre aux compagnies de pouvoir déterminer elles-mêmes leurs émissions, sans avoir à recourir aux services de consultants spécialisés, qui coutent souvent trop cher pour les petites à moyennes entreprises. Bien entendu, la qualité et légitimité des résultats obtenus est bien en dessous des services offerts par des professionnels. Néanmoins, ce logiciel donne accès à tout business à un moyen d'analyser sa chaine de valeur et de déterminer où des efforts sont à prioriser pour réduire son empreinte carbone. Le BEE s'appuie sur EXIOBASE3 et sur une base de données hybride générée avec la version 1.0 de pyLCAIO. *Ce logiciel constitue donc une première application concrète des résultats de cette recherche.*

7.2 Régionalisation et hybridation

Lors du dernier article de cette thèse, la base de données EXIOBASE a été liée à la méthodologie d'impact IMPACT World+ (IW+), afin de pouvoir étudier l'effet de l'hybridation sur 13 catégories d'impact. Une des particularités de la méthode IW+ est la possibilité d'utiliser des facteurs de caractérisation régionalisés, on parle de régionalisation des impacts. Cette régionalisation permet de refléter le fait que pour certaines catégories d'impact, un polluant ne va pas affecter la catégorie d'impact avec la même force, dépendamment des facteurs géographiques. Ainsi, pour une zone A au sol beaucoup plus fragile que le sol de la zone B, le polluant aura un impact plus élevé s'il est émis dans la zone A.

La régionalisation des impacts nécessite une spatialisation de l'inventaire, i.e., déterminer dans quelle région le polluant est émis. Cette spatialisation est compliquée en ACV, comme beaucoup de processus élémentaires sont définis pour des régions très étendues (e.g. Europe). En IO (et plus particulièrement dans EXIOBASE), cette spatialisation est assurée au niveau national. J'ai donc implémenté la régionalisation des impacts pour EXIOBASE, afin de comparer l'augmentation (ou diminution) apportée par la régionalisation des impacts à l'augmentation due à l'hybridation.

Malheureusement, il semble y avoir une inconsistance pour certaines catégories d'impact de IW+. La figure 7-1 montre l'augmentation/diminution de l'impact après régionalisation pour les impacts du cycle de vie de la consommation des 49 pays d'EXIOBASE, chaque pays étant une ligne. On observe que pour l'eutrophisation de l'eau fraiche (en violet), la régionalisation tend fortement à baisser les impacts du cycle de vie sur cette catégorie pour la grande majorité des pays. De façon similaire, l'acidification des terres (orange) et l'occupation des terres (bleu) tendent à être majoritairement augmentées par la régionalisation. La version non régionalisée des facteurs de caractérisation représentant un impact moyen, on s'attendrait à avoir une distribution plus proche de celles de la rareté de l'eau (vert), de l'acidification de l'eau fraiche (rouge) et de l'eutrophisation de l'eau marine (marron) où la régionalisation entraine à la fois des augmentations et diminutions pour différentes zones géographiques.

Cette première application des facteurs régionalisés d'IW+ à un système complet, i.e., EXIOBASE permet de modéliser l'économie entière, a montré qu'il y a probablement un problème dans l'agrégation de données de IW+, utilisée pour générer les facteurs de caractérisation régionalisés.

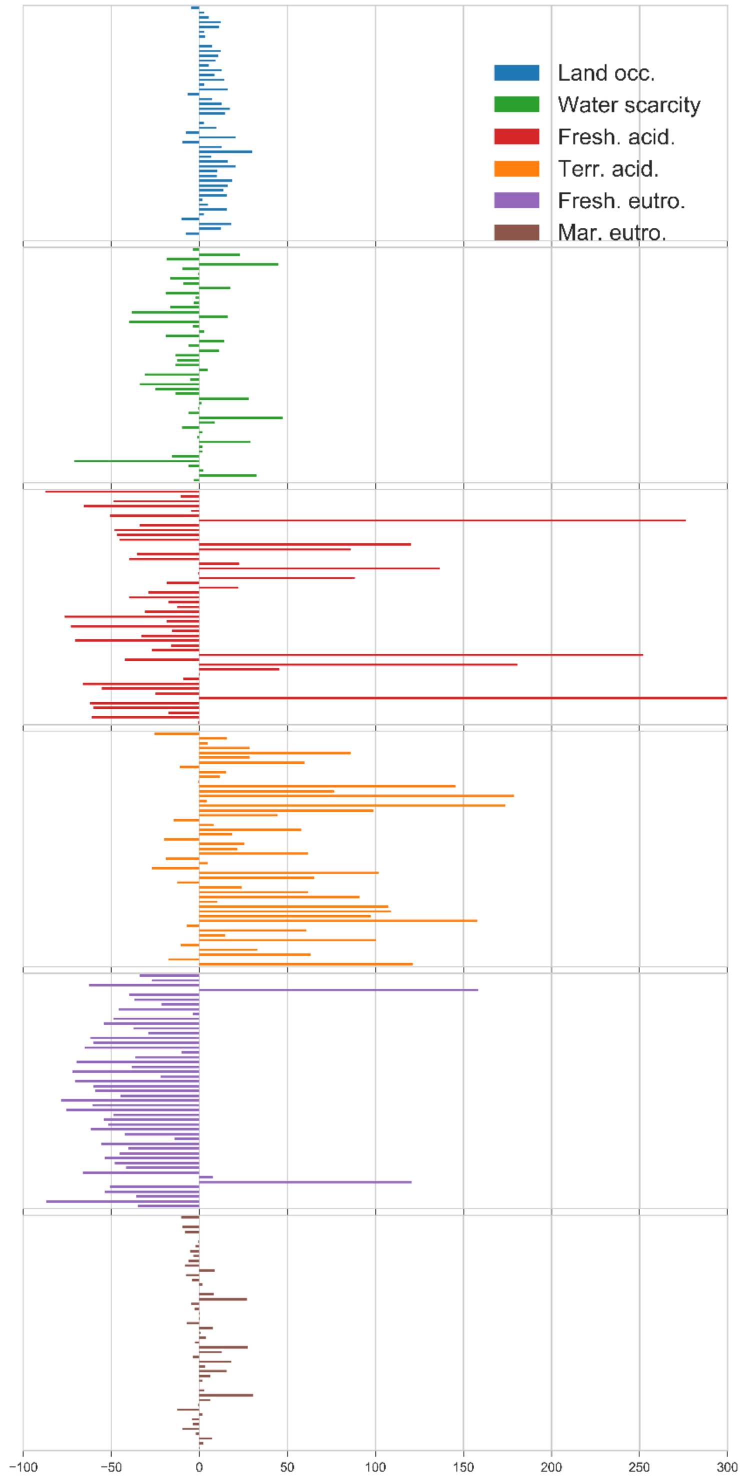


Figure 7-1 Premiers résultats de régionalisation des impacts avec EXIOBASE. Chaque rangée représente les impacts du cycle de vie de la consommation d’un pays. L’axe des abscisses montre l’augmentation/diminution relative de l’impact après régionalisation.

7.3 Contributions à d'autres articles scientifiques

Les résultats de cette recherche ont été appliqués dans l'article suivant :

Hung, C. R., Völler, S., Agez, M., Majeau-Bettez, G., & Strømman, A. H. (2020). Regionalized footprints of battery electric vehicles in Europe. *Environmental Research Letters* [in review].

7.4 Participation aux conférences

Ce projet de recherche a été présenté dans les conférences suivantes :

Agez, M., Majeau-Bettez, G., Margni, M., Strømman, A. H., Samson, R. (2019). The streamlined hybridization of Life Cycle Assessment and Environmentally Extended Input-Output databases, application to Ecoinvent and Exiobase. In *2019 International Society for Industrial Ecology*. Beijing, China.

CHAPITRE 8 DISCUSSION GÉNÉRALE

8.1 Atteinte des objectifs fixés

- **Développer une méthode de correction du double comptage en ACVH semi-automatique**

La revue de littérature de l'article 1 a permis d'identifier plusieurs méthodes existantes, dont une seule pouvait être mise en place de façon semi-automatique (binary method). En revanche, elle était limitée dans la performance du double comptage, en laissant beaucoup d'incidents de double comptage intouchés, voire en laissant des incohérences technologiques (e.g., intrants d'essence dans l'usage d'une voiture électrique). Une nouvelle méthode de correction pour le double comptage a donc été développée (STAM) dans ce travail de recherche basée sur la méthode déjà existante identifiée, avec le but d'améliorer la performance de cette méthode.

Au final, STAM a été appliquée de façon automatique à plus de 6,000 procédés. Les résultats obtenus avec STAM ne contiennent que quelques valeurs aberrantes, qui pourraient même ne pas être reliées à la correction du double comptage. STAM est aussi adaptable et transparente. L'utilisateur souhaitant modifier les hypothèses choisies dans cette recherche peut simplement éditer un fichier Excel qui sera ensuite lu et utilisé par pyLCAIO pour générer une nouvelle version de la base de données hybride.

- **Développer une base de données hybride pour la réalisation d'études ACVH comparatives**

Une base de données hybride a bien été développée et présentée dans l'article 2 de cette thèse et a été améliorée davantage avec l'article 3. Cette base de données peut d'ores et déjà être utilisée pour réaliser des études comparatives. Cependant, cela requiert une connaissance de la manipulation de données matricielles. En d'autres mots, cette base de données hybride ne peut actuellement pas être importée et réutilisée telle quel dans les logiciels ACV déjà existants. Il faudrait que la base de données obtenue soit lisible par les logiciels ACV qui adoptent tous un format de données différent.

La base de données générée a également fourni un ordre de grandeur représentatif du niveau de troncature de l'ACV, et ce sur plusieurs catégories d'impact et plusieurs catégories de produit (voir figure 6-2). Ainsi, on peut voir que la troncature en ACV est de l'ordre de 10% pour l'indicateur de réchauffement climatique. Par conséquent, toute ACV non hybride typiquement sous-estime les impacts sur le réchauffement climatique de 10%. L'estimation de ce niveau de troncature sur plusieurs indicateurs environnementaux est également une première. En effet, il apparait que toutes les catégories d'impact ne sont pas affectées de la même manière par la troncature en ACV. Ainsi, l'occupation des terres est typiquement sous-estimée de 200% dans l'ACV. Cette recherche permet aux praticiens ACV souhaitant améliorer la performance de leurs inventaires de cycle de vie générés de prioriser les efforts à apporter.

- **Développer une méthode d'hybridation ne dépendant pas du prix du produit de référence**

Comme exposé dans la revue de littérature de ce mémoire et dans l'article 3, la donnée de prix est souvent limitante en ACVH. Par exemple, la base de données hybride générée dans l'article 2 était incapable d'hybrider la plupart des processus de traitement de déchets, puisque ceux-ci n'ont pas de prix associés dans ecoinvent. Cette recherche a donc abouti à une autre façon d'hybrider ces procédés. En l'occurrence, en n'utilisant non pas le prix du flux de référence, mais le prix d'un flux commun à tout produit du secteur IO servant à l'hybridation. Le flux commun doit également être bien quantifié en ACV, car il est considéré non tronqué dans cette nouvelle méthode d'hybridation. Cette nouvelle façon d'hybrider pourrait également avoir un impact conséquent sur l'IO en général étant donné les efforts croissants des agences statistiques pour capturer au moins certains flux en unités physiques.

Cette méthode a permis l'hybridation de 982 procédés supplémentaires de la base de données ecoinvent. Parmi ces procédés en revanche, il y a des valeurs aberrantes, atteignant les 1000% d'augmentation après hybridation, et ce sur plusieurs catégories d'impact. Ce premier essai de la méthode requiert donc d'être raffiné et testé davantage avant de pouvoir potentiellement supplanter la méthodologie classique de l'ACVH.

- **Intégrer l'endogénéisation des capitaux et la génération de flux élémentaires additionnels dans le cadre d'hybridation développé**

Les travaux de Södersten sur l'endogénéisation des capitaux d'EXIOBASE et les travaux de Muller sur l'addition d'extensions environnementales à EXIOBASE ont été intégrés avec succès dans le cadre d'hybridation développé dans l'article 2. L'impact respectif de ces deux mesures correctives supplémentaires a également été identifié et représenté dans la figure 6-2 de l'article 3. On y voit que l’endogénéisation des capitaux a un impact non négligeable sur les résultats obtenus, alors que l’addition d’extensions environnementales a un effet limité. Ceci peut être dû aux problèmes rencontrés dans l’article 3 avec la catégorie « freshwater ecotoxicity » étant donné que la plupart des nouvelles émissions impactaient majoritairement cette catégorie d’impact.

8.2 Limites identifiées

8.2.1 Une hybridation *semi*-automatique

La cadre d’hybridation qui a été développé dans cette recherche correspond à un cadre semi-automatique. Autrement dit, il y a toujours des étapes manuelles à effectuer, notamment, des étapes de concordances entre bases de données. Ces étapes manuelles peuvent dans certains cas apporter plus de contrôle à l’utilisateur (comme c’est le cas avec la méthode STAM), mais dans le cas des concordances c’est une source d’erreurs potentielles (si un procédé d’ecoinvent et relié au mauvais secteur d’EXIOBASE), mais aussi un frein à la réutilisation du cadre d’hybridation. Actuellement, seules les versions spécifiques ecoinvent3.5 et EXIOBASE3 peuvent être hybridées, car seules les concordances entre ces deux versions ont été générées dans cette recherche. Un utilisateur souhaitant hybrider ecoinvent3.7 avec GTAP, par exemple, devra de lui-même générer des concordances ainsi que d’autres paramètres clés.

8.2.2 Hybridation post allocation

Le cadre d’hybridation développé dans cette recherche fonctionne post-allocation. Autrement dit, l’hybridation ne peut être opérée qu’avec une version d’ecoinvent allouée et une matrice technologique d’EXIOBASE. Ce choix a été fait pour garantir une facilité d’accès pour l’utilisateur étant donné que la version non allouée d’ecoinvent ne peut couramment être obtenue que sur demande. Le processus d’hybridation étant déjà compliqué, imposer à l’utilisateur de se procurer une version non allouée d’ecoinvent découragerait l’utilisateur.

Ce choix entraine deux problèmes en revanche. (1) Les coproduits dans ecoinvent et EXIOBASE ne sont pas alloués de la même façon. EXIOBASE alloue les co-produits en fonction de leur valeur économique, alors qu’ecoinvent majoritairement utilise l’allocation économique également, mais a recours à d’autres types d’allocations régulièrement. Cela crée des incohérences dont il faudrait mesurer l’impact sur les résultats obtenus. (2) L’hybridation présentée dans cette recherche se base sur les prix fournis par la base de données ecoinvent. Ces prix toutefois, ont été compilés avant allocation. Pour les co-produits de ecoinvent, la donnée de prix peut ne pas être adaptée, car le processus d'allocation ne respecte pas toujours l'équilibre monétaire. Si l'hybridation était faite préallocation, ce problème disparaitrait. Dans les cas extrêmes où le prix d'un produit devient inférieur à la somme des prix des intrants, les procédés n'ont pas été hybridés (dans l'article 2). Ces procédés ont ensuite été hybridés (dans l'article 3) en utilisant la méthode d'hybridation ne dépendant pas du prix du produit de référence.

8.2.3 Incertitudes

Le but de ce projet de recherche était d'améliorer la qualité des données d'arrière-plan en ACV. Cependant, cette recherche n'inclut aucun travail sur l'incertitude. Il est possible que la somme de toutes les incertitudes introduites par ce projet dépasse l'erreur de troncature de l'ACV, limitant fortement le bénéfice principal apporté par cette recherche. Il y a quatre incertitudes notoires dans ce projet de recherche qui mériteraient d'être quantifiées.

- Les incertitudes en IO ne sont pas quantifiées, contrairement à l'ACV. En revanche, elles sont principalement corrélées au biais d'agrégation de l'IO qui peut être estimé en comparant données agrégées et désagrégées d'une base de données IO, dans le cas de cette recherche : EXIOBASE. Il serait possible d'agréger certains secteurs similaires (e.g., tous les types d'électricité ou tous les produits issus de la raffinerie) d'EXIOBASE et de comparer les résultats obtenus afin de dégager une variance qui pourrait servir d'estimation cette incertitude.
- Ecoinvent compile des données de prix (utilisées dans ce projet de recherche) mais précise dans sa documentation que ces données sont très incertaines. Les intrants ajoutés par l'ACVH au processus élémentaire de l'ACV ont une relation linéaire avec la donnée de prix utilisée. L'incertitude autour de ces prix pourrait être estimée en regardant la variabilité des prix utilisés comparés aux prix de bases de données spécialisées tel le BACI qui est tiré de Comtrade.
- La correction du double comptage a été automatisée dans cette recherche. Cela a permis son application à plusieurs milliers de procédés, chose qui aurait été impossible si cette correction avec été faite manuellement. Cependant, automatiser cette correction entraine

des imprécisions. Certains intrants pourraient être retirés de l'inventaire hybride alors qu'ils ne devraient pas l'être et au contraire, certains intrants pourraient ne pas être retirés alors qu'ils créent un double comptage. Pour estimer cette incertitude, on pourrait utiliser la variation des résultats entre les différentes méthodes de double comptage. Yu et Wiedmann (2018) implémentent une stratégie similaire dans leur article.

- Ce projet de recherche a nécessité plusieurs concordances : entre les procédés d'ecoinvent et les secteurs d'EXIOBASE, entre les flux élémentaires de USEEIO et de EXIOBASE et finalement entre les secteurs de USEEIO et EXIOBASE. Ces concordances ont été faites manuellement et elles sont donc susceptibles d'être fausses à certains points. Soit par erreur d'inattention, soit parce que le nom du secteur/procédé n'est pas ce qu'il parait être. Cette incertitude pourrait être estimée en associant volontairement des procédés au mauvais secteur et réalisant une moyenne de la variation observée.

8.3 Développements futurs potentiels

Hormis les limites identifiées à la section précédente qui mériteraient d'être améliorées, il y a également des pistes de recherche dégagées de ce projet de recherche.

8.3.1 Intégration dans des logiciels existants

L'une des principales limites de l'ACVH et sa complexité inhérente. Même si ce projet a simplifié le processus d'hybridation en l'automatisant et en réduisant la quantité de données à manipuler par l'utilisateur, cela reste complexe à utiliser, notamment car cela requiert de savoir manipuler des matrices de plusieurs millions d'entrées ainsi que de savoir utiliser une librairie Python. Afin de pouvoir rendre cette recherche abordable et d'encourager son adoption par les praticiens ACV, une version de la base de données hybride devrait être accessible dans les logiciels ACV couramment utilisés, i.e., SimaPro, openLCA et brightway2. Cependant, seul brightway2 est open source parmi ces logiciels. L'intégration de ce projet de recherche et d'une base de données hybride utilisable dans brightway2 devrait donc être priorisée. Cela requiert la création d'un module importateur à intégrer dans le code de brightway2, transformant la base de données hybride du format pandas.DataFrame au format utilisé par brightway2 : un dictionnaire dont chaque clé est une activité d'ecoinvent (i.e., une colonne du pandas.DataFrame) et chaque valeur correspond aux échanges de cette activité avec les autres activités de la base de données (i.e., chaque valeur de la colonne sélectionnée).

8.3.2 Amélioration des facteurs de caractérisation régionalisés d'IW+ grâce à l'IO

Il y a deux composants d'un facteur de caractérisation régionalisé : un facteur dit natif et une quantité de flux annuelle (aussi appelée proxy par IW+) (Bulle et al., 2019). Les facteurs de caractérisation natifs sont fournis à une échelle bien plus fine que l'échelle du pays qui est souvent utilisée en ACV et en IO. Ces facteurs natifs sont donc agrégés au niveau du pays ou d'une région en utilisation les proxys. Les proxys utilisés dans IW+ cependant, ne sont toujours pas adaptés. Par exemple, les facteurs natifs de caractérisation pour l'eutrophisation d'eau douce sont agrégés en utilisant la densité de population. Cela suppose que plus il y a de population à un endroit, plus l'eutrophisation sera importante à ce même endroit. Une meilleure donnée d'agrégation pour cette catégorie d'impact serait l'activité agricole. Cette donnée peut être trouvée dans les bases de données IO globales (pour pouvoir agréger les facteurs nationaux en facteurs continentaux) et régionales (pour pouvoir agréger les facteurs natifs en facteurs nationaux). L'IO pourrait donc servir à améliorer la détermination des facteurs de caractérisation régionalisés en fournissant des meilleurs proxys.

8.3.3 Désagrégation de secteurs IO grâce à l'ACV

Le cadre d'hybridation présenté dans cette recherche cherche à compléter les processus élémentaires de l'ACV en utilisant les données de l'IO. L'IO ne tire donc aucune plus-value *directe* de ce projet. En d'autres termes, les données de l'IO n'ont pas été améliorées directement. Il est possible toutefois de désagréger des secteurs IO grâce à des données ACV (et donc grâce aux données d'ecoinvent). Mathématiquement, cela revient à compléter les données de la matrice C^d de l'équation (6) du premier article, jusqu'à présent fixées à zéro dans cette recherche. Conceptuellement, cette matrice représente l'utilisation des processus ACV par l'économie. Par quels secteurs de l'économie ces procédés sont-ils achetés ? Une fois un procédé ACV relié au reste de l'économie, il peut ensuite être désagrégé du secteur auquel il se rapportait initialement dans l'IO. Par exemple, on pourrait intégrer le processus ACV « production de voiture électrique » et relier ce procédé au reste de l'économie (e.g., ce procédé est demandé par les concessionnaires automobiles, par les consommateurs, etc.) pour ensuite pouvoir désagréger le secteur économique « voitures » en « voitures thermiques » et « voitures électriques ». Bien entendu le nouveau secteur « voitures électriques » serait modélisé par l'ACV.

Avec cette désagrégation, on vient apporter de la spécificité à l'IO. Pour continuer avec l'exemple précédent, l'IO serait maintenant en mesure de modéliser l'impact sur l'environnement de voitures électriques, ce qui pourrait mener à des études systémiques de l'implantation de voitures électriques sur un territoire donné (pays, continent, monde). Dans cette optique, les données ACV et IO se complètent et sont intégrées ensemble, d'où le nom d'ACVH intégrées.

La désagrégation n'est toutefois pas triviale, car elle requiert une quantité importante de données supplémentaires sur l'utilisation par l'économie du/des produit(s) à désagréger. Il faut donc être en mesure de renseigner précisément cette utilisation. Se baser uniquement sur une répartition économique des produits à désagréger (i.e., 25% d'intrants du secteur « voitures » sont associés à « voitures électriques » parce que les voitures électriques représentent 25% des achats de voitures) va mener aux mêmes résultats qu'une IO non désagrégée.

8.3.4 Meilleure estimation des capitaux en ACV

Les capitaux représentent une grande part des émissions mondiales (24%) et pourtant leur couverture en ACV est limitée. Les quelques capitaux présents dans les bases de données ACV type ecoinvent (e.g., bâtiments, machines, véhicules), sont souvent représentés par seulement une poignée de modélisations originales qui sont ensuite réutilisées pour d'autres capitaux similaires. De plus, certains capitaux sont absents de ces bases de données. L'IO en revanche, couvre de façon plus exhaustive et spécifique les capitaux disponibles. Les générateurs de données ACV pourraient donc s'appuyer sur l'IO pour générer de meilleures données de capitaux pour l'ACV. Cela sous-entend recourir à une version endogénéisée d'une base de données IO et que la base de données IO sélectionnée couvre plusieurs années pour pouvoir faire une moyenne. Utiliser une année donnée induira des erreurs étant donné que la construction de capital n'a pas lieu de façon constante sur une année, mais a bien lieu sur plusieurs années.

8.4 Recommandations

Au vu de l'apport de cette recherche, un nombre de recommandations peut être effectué pour les différentes parties prenantes de l'ACV, l'IO et l'ACVH.

8.4.1 Pour les praticiens

- Cette thèse a apporté un chiffre représentatif de la troncature de l'ACV classique, et ce sur plusieurs indicateurs environnementaux. A la vue de ces résultats, l'application de l'ACVH devrait être priorisée par les praticiens. C'est plus particulièrement vrai pour les catégories d'impact comme l'occupation des terres ou l'eutrophisation de l'eau marine, qui moyenne des sous-estimations de l'ordre de 200% et 75% respectivement. L'hybridation devrait idéalement inclure une version endogénéisée de la table IO utilisée, étant donné l'impact de l'endogénéisation sur les résultats de la figure 6-2.
- Le premier article de ce travail de recherche a démontré la grande variabilité des résultats donnés par un inventaire hybride suite au choix de telle ou telle méthode de correction pour le double comptage. Afin d'assurer une transparence, une reproductibilité et une pertinence de l'étude, il est important que les études ACVH publiées prennent le temps de détailler la méthode de correction pour le double comptage qui a été appliquée.

8.4.2 Pour les fournisseurs de données/bases de données

- Le problème majeur identifié dans l'hybridation des bases de données ecoinvent et EXIOBASE a été le manque de distinction entre une donnée nulle et une donnée absente, i.e., entre de vrais et faux zéros. Actuellement, en compilation de données ACV, une donnée qu'on sait nulle n'est pas renseignée comme telle et finit donc au même statut qu'une donnée absente. Pour l'ACVH, qui cherche à compléter uniquement les données absentes, cette distinction est primordiale. Sans elle, les praticiens ACVH sont réduits à devoir « deviner » quelles données sont manquantes et quelles données sont réellement non requises par un procédé. Ce fut notamment le cas dans l'application de la méthode STAM et c'est également la principale limite de la méthode binaire de correction pour le double comptage. Pour une application efficace sur le terrain ne requérant que peu d'efforts supplémentaires, cela demanderait des formulaires composés de secteurs de l'économie (basés sur les classifications NACE, NAICS ou ISIC par exemple) sur lesquels il faudrait renseigner l'achat ou non de produits de ces secteurs. La quantité achetée n'importerait pas. Le but avec ce formulaire serait uniquement de distinguer les vrais des faux zéros. Les faux zéros, i.e., les données que le fournisseur de données sait manquantes, seraient ensuite estimées en utilisant l'hybridation.

- Comme suggéré dans les pistes de recherche futures identifiées dans cette thèse, les compilateurs de bases de données ACV pourraient extraire de meilleures données sur les capitaux de bases de données IO où les capitaux sont endogénéisés. Similairement, les procédés « market » de bases de données ACV tel que ecoinvent sont, au final, une tentative de représenter la distribution des produits. Cette distribution pourrait également être améliorée basé sur les données brutes de l'IO qui détaillent beaucoup mieux les étapes de distribution.
- Spécifiquement pour la base de données ecoinvent, actuellement, l'identifiant (UUID) de chaque procédé change entre chaque version. Les concordances développées pour hybrider ecoinvent et EXIOBASE deviennent donc obsolètes à chaque mise à jour d'ecoinvent, étant en partie basées sur l'identifiant du procédé.
- Les compilateurs de bases de données IO pourraient fournir des extensions environnementales spatialisées (i.e., associer un lieu d'émission à une émission). Cela faciliterait l'application de facteurs de caractérisation régionalisés.

8.4.3 Pour les communautés ACV, IO et ACVH en général

- Le fléau majeur de la combinaison de données en écologie industrielle, c'est les problèmes de classification. Ces problèmes ont constitué une partie non négligeable des défis techniques de cette recherche. Chaque partie prenante de la communauté choisit la classification qu'il ou elle préfère ou juge plus pertinent, ce qui vient ensuite compliquer les travaux de combinaison de données par la suite. Les communautés ACV, IO et MFA (Material Flow Analysis) devraient donc idéalement adopter un unique système de classification. Un système de classification indépendant, comme celui de comtrade, pourrait être choisi par exemple, et chaque ensemble de données se rattacherait à cette classification.

- Pour la communauté ACVH spécifiquement, il y a un vrai défi d'accessibilité et de compréhension de la méthodologie. La communauté ACVH n'est elle-même pas en accord sur l'application des différentes méthodologies existantes (e.g. méthode par paliers vs méthode intégrée). Comment alors les communautés ACV et IO pourraient-elles appliquer l'ACVH en toute confiance ? Il y a donc un travail d'harmonisation de la terminologie et des méthodologies d'ACVH à organiser. De plus, une fois ce travail terminé, il faudra développer des outils simples d'utilisation pour réellement populariser l'application de l'ACVH.

- L'ACV vit une crise de la donnée. On attend toujours plus de l'ACV étant donné sa popularité croissante, mais elle souffre de manque de données. L'une des raisons pour cela est la non-consolidation des acquis de l'ACV dans les bases de données. Nombre d'études ACV ne sont pas intégrées dans les bases de données ACV. Le formatage des données d'études ACV pour intégration dans une base de données est peut-être trop demandant. Une solution alternative est l'utilisation de bases de données autres (ici le terme bases de données correspond plus au terme anglophone « dataset » que le terme anglophone « database ».). Une plateforme regroupant nombre de données provenant d'études ACV, IO ou MFA a été récemment développée (Pauliuk, Heeren, Hasan, & Müller, 2019). Les communautés de l'écologie industrielle ont toutes avantage à venir peupler ce genre de bases de données alternatives.

- L'ACV vit également une crise de pertinence. Par sa méthodologie, l'ACV n'est pas pertinente au-delà de l'analyse isolée d'un changement marginal (au-delà de l'échelle de la chaine de valeur). L'ACV a donc une pertinence limitée pour les analyses de scénario de transformation. Les gouvernements et grandes multinationales sont à la recherche de réponses pour examiner des transformations profondes du métabolisme socioéconomique, et l'ACV y joue un rôle limité actuellement. L'hybridation en revanche, permettrait à l'ACV de venir se positionner sur ce terrain, notamment avec la méthodologie intégrée qui lui donnerait enfin accès à une analyse à l'échelle du système global.

- A l'heure actuelle, le praticien ACV peut être considéré comme un créateur de descriptions technologiques, récoltant données d'avant-plan et créant de nouveaux procédés à partir de celles-ci. Le praticien ACVH quant à lui joue plutôt le rôle d'un désagrégateur de chaines de valeur. Son but est de s’appuyer sur les données déjà existantes de l’IO et de désagréger ces dernières pour obtenir de nouveaux procédés. Dans un futur proche, il serait possible que tout praticien ACV s’appuie davantage sur l’IO, réduisant ainsi sa collecte de données et se rapprochant du statut d’un désagrégateur de données à son tour. Si l’on regarde encore plus loin dans le futur, lorsque les bénéfices de l’industrie 4.0 arriveront dans le domaine de l’ACV, les rôles des praticiens ACV et ACVH seront modifiés. Imaginons par exemple, l’enregistrement automatique des données de procédés et leur acheminement (e.g. via blockchain). Les praticiens ACV/ACVH n’auraient dès lors, plus de données à collecter, ou à désagréger car ces données seraient prélevées automatiquement. Ultimement, si les données de chaque activité de l’économie étaient prélevées automatiquement, il n'y aurait alors plus de distinction entre avant-plan et arrière-plan. Les praticiens ACV/ACVH passeraient donc à un rôle d’assembleur de chaines de valeurs.

CHAPITRE 9 CONCLUSION

Ce projet a permis de répondre à la question de recherche principale : comment améliorer la complétude des inventaires de cycle de vie tout en gardant la spécificité de l'ACV ? Cette complétude est idéalement atteinte grâce à l'adoption de l'ACVH et l'utilisation du cadre d'hybridation automatisé présenté dans ce mémoire. Les principales contributions de ce projet de recherche sont :

- Une étude exhaustive sur les méthodes de correction de double comptage existantes dans le domaine de l'ACVH, où les différentes méthodes sont listées, décrites, appliquées à un même exemple et leurs résultats comparés entre eux, ce qui a mis en lumière l'importance du choix de la méthode de correction du double comptage en ACVH.
- Deux nouvelles méthodes de correction du double comptage semi-automatique. La première, STAM, est plus précise que les méthodes déjà existantes ; la seconde, DSET, dont l'application et la pertinence sont intimement reliées à la couverture des technologies par les bases de données ACV. Actuellement, cette méthode n'est pas applicable. Ces deux méthodes apportent une meilleure correction des incidents de double comptage en ACVH.
- Un cadre méthodologique pour l'hybridation de milliers de procédés de façon semi-automatique et la traduction de ce cadre en logiciel (pyLCAIO). Ce logiciel permet la création de bases de données hybrides. Le cadre peut aussi être adapté pour hybrider différentes combinaisons de bases de données ACV et IO.
- Une base de données hybride utilisable pour réaliser des études ACV ou ACVH comparatives. Cette base de données hybride apporte une meilleure quantification des impacts de cycle de vie liés aux données d'arrière-plan.
- Une estimation de la troncature en ACV, basée sur un échantillon représentatif (la base de données ecoinvent dans son ensemble), pour 13 catégories d'impact différentes et ce pour 11 catégories de produit différentes. Cette estimation renseigne sur la sous-estimation typique en ACV et alerte donc sur la pertinence des résultats absolus obtenus par l'ACV.

- Une nouvelle méthode d’hybridation non dépendante du prix du produit de référence, étant donc capable d’hybrider des procédés dont le prix est soit non fiable, soit indisponible. Cette méthode d’hybridation permet l’hybridation de procédés pour lesquels la donnée de prix est considérée non fiable ou est indisponible.
- L’expansion du cadre méthodologique déjà introduit dans cette recherche pour intégrer l’endogénéisation des capitaux de l’IO. Cette intégration des capitaux de l’IO dans le cadre d’hybridation complète les résultats obtenus par l’ACVH et peut être réutilisée par la communauté.

Plusieurs limites à ce projet ont été identifiées, notamment la difficulté d’adapter le cadre d’hybridation développé à d’autres bases de données ACV et IO, ce qui est vient de la dépendance de concordances générées manuellement ; le choix d’effectuer l’hybridation post allocation, ce qui entraine des inconsistances au niveau de l’allocation et des prix utilisés lors de l’hybridation ; le fait que ce projet n’intègre aucune étude des incertitudes alors que la somme des incertitudes introduites pourrait dépasser la troncature corrigée. Certaines pistes ont été dégagées pour traiter quelques-unes de ces limites comme l’estimation des incertitudes liées aux concordances effectuées dans ce projet en faisant une moyenne de la variabilité observée dans les résultats une fois une mauvaise concordance simulée ; ou comme l’étude de la variabilité entre prix utilisés et prix venant de bases de données spécialisées pour les prix.

De plus, de nouvelles voies de recherche ont été identifiées : (1) l’amélioration des facteurs de caractérisation régionalisés grâce à l’IO, où les données d’agrégation utilisées pour agréger les facteurs de caractérisation natifs en facteurs de caractérisation régionalisés pour un pays ou une région peuvent être supplantées par des données provenant de l’IO ; (2) l’amélioration des données de capitaux en ACV qui proviendrait de moyennes agrégées sur une série temporelle de bases de données IO dans lesquelles les capitaux ont été endogénéisés ; (3) la désagrégation de certains secteurs de l’IO en utilisant les données ACV (e.g., venant de ecoinvent) afin de permettre la réalisation d’études systémiques et de modéliser grossièrement l’impact de l’adoption d’une nouvelle technologie ou d’un changement de consommation sur l’économie entière ou nationale.

Ultimement ce projet de recherche a contribué à une meilleure quantification des émissions et de leurs impacts sur l’environnement. En effet, l’ACV est de plus en plus utilisée et adoptée par les preneurs de décisions, mais la troncature, omniprésente dans l’ACV, peut éventuellement mener à des conclusions incorrectes ne donnant pas au décideur les informations dont il a besoin. Ce projet devrait donc encourager les spécialistes à adopter l'ACVH étant donnée l'estimation de la troncature qui est ressortie de ce travail. De plus, dans le cas où les ressources pour réaliser une ACVH ne sont pas disponibles, la base de données hybride développée peut d'ores et déjà permettre des ACV plus justes. Des estimations plus justes *devraient* mener à des décisions éclairées, pour un monde meilleur.

RÉFÉRENCES

Acquaye, A. A., Wiedmann, T., Feng, K., Crawford, R. H., Barrett, J., Kuylenstierna, J., … McQueen-Mason, S. (2011). Identification of "Carbon Hot-Spots" and Quantification of GHG Intensities in the Biodiesel Supply Chain Using Hybrid LCA and Structural Path Analysis. *Environmental Science & Technology*, *45*(6), 2471–2478. https://doi.org/10.1021/es103410q

ADEME. (2020). Agribalyse. Retrieved from https://ecolab.ademe.fr/agribalyse

Agez, M., Majeau-Bettez, G., Margni, M., Strømman, A. H., & Samson, R. (2019). Lifting the veil on the correction of double counting incidents in hybrid Life Cycle Assessment. *Journal of Industrial Ecology*, *24*(3), 1–17. https://doi.org/https://doi.org/10.1111/jiec.12945

Agez, M., Wood, R., Margni, M., Strømman, A. H., Samson, R., & Majeau-Bettez, G. (2020). Hybridization of complete LCA and MRIO databases for a comprehensive product system coverage. *Journal of Industrial Ecology*, 1–17. https://doi.org/10.1111/jiec.12979

Andrew, R. M., & Peters, G. P. (2013). A multi-region input-output table based on the global trade analysis project database (GTAP-MRIO). *Economic Systems Research*, *25*, 99–121.

Arvesen, A., Birkeland, C., & Hertwich, E. G. (2013). The importance of ships and spare parts in LCAs of offshore wind power. *Environmental Science and Technology*, *47*(6), 2948–2956. https://doi.org/10.1021/es304509r

Arvesen, A., & Hertwich, E. G. (2011). Environmental Implications of large-scale adoption of wind power: a scenario-based life cycle assessment. *Environ. Res. Lett.*, *6*(4).

Arvesen, A., Nes, R. N., Hertwich, E. G., & Huertas-Hernando, D. (2014). Life cycle assessment of an offshore grid interconnecting wind farms and customers across the North Sea. *The International Journal of Life Cycle Assessment*, 826–837.

Asger Olsen, J. (2000). Aggregation in macroeconomic models: an empirical input–output approach. *Economic Modelling*, *17*(4), 545–558.

Azapagic, A., & Clift, R. (1998). Linear programming as a tool in life cycle assessment. *International Journal of Life Cycle Assessment*, *3*(6), 305–316.

https://doi.org/10.1007/BF02979340

Blonk consultants. (2017). Agri-footprint. Retrieved from https://www.agri-footprint.com/

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

Bullard, C. W., & Penner, P. S. (1978). Net Energy Analysis: Handbook for Combining Process and Input-Output Analysis. *Resources Energy*, *267–313*.

Bulle, C., Margni, M., Patouillard, L., Boulay, A. M., Bourgault, G., De Bruille, V., … Jolliet, O. (2019). IMPACT World+: a globally regionalized life cycle impact assessment method. *International Journal of Life Cycle Assessment*, 1653–1674. https://doi.org/10.1007/s11367-019-01583-0

Burfisher, M. E. (2011). *Introduction to Computable General Equilibrium Models*. Cambridge University Press.

Chen, Z. M., Ohshita, S., Lenzen, M., Wiedmann, T., Jiborn, M., Chen, B., … Liu, Z. (2018). Consumption-based greenhouse gas emissions accounting with capital stock change highlights dynamics of fast-developing countries. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05905-y

Ciroth, A., Muller, S., Weidema, B., & Lesage, P. (2016). Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *International Journal of Life Cycle Assessment*, *21*(9), 1338–1348. https://doi.org/10.1007/s11367-013-0670-5

Crawford, R. H. (2009). Life cycle energy and greenhouse emissions analysis of wind turbines and the effect of size on energy yield. *Renewable and Sustainable Energy Reviews*, *13*(9), 2653–2660. https://doi.org/10.1016/j.rser.2009.07.008

Crawford, R. H., Bontinck, P.-A., Stephan, A., Wiedmann, T., & Yu, M. (2018). Hybrid life cycle inventory methods – A review. *Journal of Cleaner Production*, *172*, 1273–1288. https://doi.org/10.1016/j.jclepro.2017.10.176

Dantzig, G. (1951). Application of the simplex method to a transportation problem. In T. C. K. (Ed.) (Ed.), *Activity Analysis of Production and Allocation* (pp. 339–347).

Defourny, J., & Thorbecke, E. (1984). Structural Path Analysis and Multiplier Decomposition within a Social Accounting Matrix Framework. *The Economic Journal*, *94*(373), 111–136. https://doi.org/10.2307/2232220

Deng, L., Babbitt, C. W., & Williams, E. D. (2011). Economic-balance hybrid LCA extended with uncertainty analysis: Case study of a laptop computer. *Journal of Cleaner Production*, *19*(11), 1198–1206. https://doi.org/10.1016/j.jclepro.2011.03.004

Duchin, F. (1992). Industrial input-output analysis: implications for industrial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(3), 851–855.

Duchin, F. (2005). A world trade model based on comparative advantage with m regions, n goods, and k factors. *Economic Systems Research*, *17*(2), 141–162. https://doi.org/10.1080/09535310500114903

Duchin, F., & Levine, S. H. (2011). Sectors may use multiple technologies simultaneously: The rectangular choice-of-technology model with binding factor constraints. *Economic Systems Research*, *23*(3), 281–302. https://doi.org/10.1080/09535314.2011.571238

Duchin, F., Levine, S. H., & Strømman, A. H. (2016). Combining Multiregional Input-Output Analysis with a World Trade Model for Evaluating Scenarios for Sustainable Use of Global Resources, Part I: Conceptual Framework. *Journal of Industrial Ecology*, *20*(4), 775–782. https://doi.org/10.1111/jiec.12303

Earth Overshoot Day. (2019). Past Earth overshoot days. Retrieved from https://www.overshootday.org/newsroom/past-earth-overshoot-days/

Ecoinvent centre. (2018). Ecoinvent database (version 3.5). [Cut-off system model]. Retrieved from the ecospold files. Retrieved December 18, 2018, from https://v35.ecoquery.ecoinvent.org/File/Files

European Commission. (2008). *Eurostat Manual of Supply, Use and Input-Output Tables*. https://doi.org/10.1016/j.physa.2013.07.063

European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, & World Bank. (2009). *System of National Accounts, 2008*. New York.

EXIOBASE Consortium. (2017). EXIOBASE (Version 3) [product-by-product industry-technology-assumption coefficients, for year 2011]. Retrieved from EXIOBASE. Retrieved March 14, 2019, from https://www.exiobase.eu/index.php/data-download/exiobase3mon?limit=20&limitstart=20

Fernández-Dacosta, C., Van Der Spek, M., Hung, C. R., Oregionni, G. D., Skagestad, R., Parihar, P., … Ramirez, A. (2017). Prospective techno-economic and environmental assessment of carbon capture at a refinery and CO2 utilisation in polyol synthesis. *Journal of CO2 Utilization*, *21*, 405–422. https://doi.org/10.1016/j.jcou.2017.08.005

Ferrao, P., & Nhambiu, J. (2009). A Comparison Between Conventional LCA and Hybrid EIO-LCA: Analyzing Crystal Giftware Contribution to Global Warming Potential. In *Handbook of Input-Output Economics in Industrial Ecology* (pp. 219–230). Springer.

Gabi solutions. (2019). GaBi Databases. Retrieved from http://www.gabi-software.com/databases/gabi-databases/

Gallego, B., & Lenzen, M. (2005). A consistent input–output formulation of shared producer and consumer responsibility. *Economic Systems Research*, *17*(4), 365–391. https://doi.org/10.1080/09535310500283492

Gibon, T., & Schaubroeck, T. (2017). Lifting the fog on characteristics and limitations of hybrid LCA. *International Journal of Life Cycle Assessment*, *22*(6), 1005–1008. https://doi.org/10.1007/s11367-017-1291-1

Gibon, T., Wood, R., Arvesen, A., Bergesen, J. D., Suh, S., & Hertwich, E. G. (2015). A Methodology for Integrated, Multiregional Life Cycle Assessment Scenarios under Large-Scale Technological Change. *Environmental Science and Technology*, *49*(18), 11218–11226. https://doi.org/10.1021/acs.est.5b01558

Hawkins, T. R., Singh, B., Majeau-bettez, G., & Strømman, A. H. (2012). Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles, *17*(1), 53–64. https://doi.org/10.1111/j.1530-9290.2012.00532.x

Heijungs, R., & Suh, S. (2002). *The computational structure of life cycle assessment.* Dordrecht: The Netherlands: Kluwer Academic Publisher.

International Organization for Standardization. (2006). *ISO 14040 - Environmental management* -

Life cycle assessment-Principles and framework.

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

Islam, S., Ponnambalam, S. G., & Lam, H. L. (2016). Review on life cycle inventory: methods, examples and applications. *Journal of Cleaner Production*, *136*, 266–278. https://doi.org/10.1016/j.jclepro.2016.05.144

Jeswani, H. K., Azapagic, A., Schepelmann, P., & Ritthoff, M. (2010). Options for broadening and deepening the LCA approaches. *Journal of Cleaner Production*, *18*(2), 120–127. https://doi.org/10.1016/j.jclepro.2009.09.023

Joshi, S. (1999). Product environmental life-cycle assessment using input-output techniques. *Journal of Industrial Ecology*, *3*, 95–120.

Junnila, S. I. (2006). Empirical comparison of process and economic input-output life cycle assessment in service industries. *Environmental Science and Technology*, *40*(22), 7070–7076.

Katelhön, A., Bardow, A., & Suh, S. (2016). Stochastic technology choice model for consequential life cycle assessment. *Environmental Science and Technology*, *50*(23), 12575–12583. https://doi.org/10.1021/acs.est.6b04270

Koesler, S., & Schymura, M. (2015). Substitution Elasticities in a Constant Elasticity of Substitution Framework – Empirical Estimates Using Nonlinear Least Squares. *Economic Systems Research*, *27*(1), 101–121. https://doi.org/10.1080/09535314.2014.926266

Lenzen, M. (2000). Errors in Conventional and Input-Output—based Life Cycle Inventories. *Journal of Industrial Ecology*, *4*(4), 127–148. https://doi.org/10.1162/10881980052541981

Lenzen, M., & Crawford, R. (2009). The path exchange method for hybrid LCA. *Environmental Science and Technology*, *43*(21), 8251–8256. https://doi.org/10.1021/es902090z

Lenzen, M., & Dey, C. (2000). Truncation error in embodied energy analyses of basic iron and steel products. *Energy*, *25*(6), 577–585. https://doi.org/10.1016/S0360-5442(99)00088-2

Lenzen, M., Moran, D., Kanemoto, K., & Geschke, A. (2013). Building Eora: a global multi-region input-output database at high country and sector resolution. *Economic Systems Research*, *25*, 20–49.

Lenzen, M., & Treloar, G. (2003). Differential Convergence of Life-Cycle Inventories toward Upstream Production Layers. *Journal of Industrial Ecology*, *6*(3), 137–160. https://doi.org/10.1162/108819802766269575

Lenzen, M., & Treloar, G. (2004). Endogenising capital - A comparison of two methods. *Journal of Applied Input-Output Analysis*, *10*, 1–11.

Leontief, W. (1936). Quantitative Input and Output Relations in the Economic Systems of the United States. *The Review of Economics and Statistics*, *18*(3), 105–125. https://doi.org/10.2307/1927837

Leontief, W. (1970). Environmental Repercussions and the Economic Strcture: An Input-Output Approach. *The Review of Economics and Statistics*, *52*, 262–271.

Lin, J., Liu, Y., Meng, F., Cui, S., & Xu, L. (2013). Using hybrid method to evaluate carbon footprint of Xiamen City, China. *Energy Policy*, *58*, 220–227. https://doi.org/10.1016/j.enpol.2013.03.007

Majeau-Bettez, G., Agez, M., Wood, R., Södersten, C., Margni, M., Strømman, A. H., & Samson, R. (2017). Streamlined Hybridization software: merging Ecoinvent and Exiobase. In *Biennial Conference of the International Society for Industrial Ecology*.

Majeau-Bettez, G., & Maxime, D. (2015). *The Input Output method and its use in LCA*. Retrieved from https://www.scorelca.org/espace_mb/etudes/SCORE-2014-04-synthang-34p2n31u8e-14.pdf

Majeau-Bettez, G., Strømman, A. H., & Hertwich, E. G. (2011). Evaluation of process- and Input-Output-based Life CycleInventory Databases with Regards to Truncation and Aggregation Issues. *Environmental Science & Technology*, *45*(23), 10170–10177. https://doi.org/10.1021/es201308x

Majeau-Bettez, G., Wood, R., & Strømman, A. H. (2014). Unified theory of allocations and constructs in life cycle assessment and input-output analysis. *Journal of Industrial Ecology*, *18*(5), 747–770. https://doi.org/10.1111/jiec.12142

Merciai, S., & Schmidt, J. H. (2017). Methodology for the construction of global multi-regional hybrid supply and use tables for the EXIOBASE v3 database. *Journal of Industrial Ecology*, *00*(0), 1–16. https://doi.org/10.1111/jiec.12713

Michelsen, O., Solli, C., & Strømman, A. H. (2008). Environmental impact and added value in forestry operations in Norway. *Journal of Industrial Ecology*, *12*(1), 69–81. https://doi.org/10.1111/j.1530-9290.2008.00008.x

Miller, R. E., & Blair, P. D. (2009a). *Input–Output Analysis. Foundations and Extensions*. Cambridge: Cambridge University Press.

Miller, R. E., & Blair, P. D. (2009b). *Input – Output Analysis*. Cambridge University Press.

Miller, T. R., Berrill, P., Wolfram, P., Wang, R., Kim, Y., Zheng, X., & Hertwich, E. G. (2019). Method for endogenizing capital in the United States Environmentally-Extended Input-Output model. *Journal of Industrial Ecology*, 1–15. https://doi.org/10.1111/jiec.12931

Mongelli, I., Suh, S., & Huppes, G. (2005). A Structure Comparison of two Approaches to LCA Inventory Data, Based on the MIET and ETH Databases. *The International Journal of Life Cycle Assessment*, *10*(5), 317–324. https://doi.org/10.1065/lca2004.12.198

Morimoto, Y. (1970). On Aggregation Problems in Input-Output Analysis. *The Review of Economic Studies*, *37*(1), 119–126. https://doi.org/10.2307/2296502

Moskowitz, P. D., & Rowe, M. D. (1985). A comparison of input-output and process analysis. *Health and Environmental Risk Assessment*, 281–293.

Muller, E. (2019). *Estimation et complétion des flux élémentaires manquants du tableau entrées-sorties multirégional EXIOBASE*. Polytechnique Montreal. Retrieved from https://publications.polymtl.ca/3920/

Muller, S., Lesage, P., & Samson, R. (2016). Giving a scientific basis for uncertainty factors used in global life cycle inventory databases: an algorithm to update factors using new information. *International Journal of Life Cycle Assessment*, *21*(8), 1185–1196. https://doi.org/10.1007/s11367-016-1098-5

Nakamura, S., & Kondo, Y. (2002). Input-output analysis of waste management. *Journal of Industrial Ecology*, *6*(1), 39–63. https://doi.org/10.1162/108819802320971632

Norris, G. A. (2002). Life Cycle Emission Distributions Within the Economy: Implications for Life

Cycle Impact Assessment. *Risk Analysis*, *22*(5), 919–930. https://doi.org/10.1111/1539-6924.00261

Patouillard, L. (2019). IMPACT World+ [midpoint1.28 global default values] retrived from https://zenodo.org/record/3521034#.XtUUBcB7lHM. Retrieved from https://zenodo.org/record/3521034#.XtUUBcB7lHM

Patouillard, L., Bulle, C., Querleu, C., Maxime, D., Osset, P., & Margni, M. (2018). Critical review and practical recommendations to integrate the spatial dimension into life cycle assessment. *Journal of Cleaner Production*, *177*, 398–412. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.12.192

Pauliuk, S., Heeren, N., Hasan, M. M., & Müller, D. B. (2019). A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype. *Journal of Industrial Ecology*, *23*(5), 1016–1027. https://doi.org/10.1111/jiec.12890

Pauliuk, S., Majeau-bettez, G., Muller, D. B., & Hertwich, E. G. (2016). Toward a Practical Ontology for Socioeconomic Metabolism. *Journal of Industrial Ecology*, *20*(6), 1260–1272. https://doi.org/10.1111/jiec.12386

Pauliuk, S., Majeau-bettez, G., Mutel, C. L., & Steubing, B. (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, *19*(6). https://doi.org/10.1111/jiec.12316

Perkins, J., & Suh, S. (2019). Uncertainty Implications of Hybrid Approach in LCA: Precision versus Accuracy. *Environmental Science & Technology*, *53*(7), 3681–3688. https://doi.org/10.1021/acs.est.9b00084

Pomponi, F., & Lenzen, M. (2017). Hybrid life cycle assessment (LCA) will likely yield more accurate results than process-based LCA. *Journal of Cleaner Production*, *150*, 237–242. https://doi.org/10.1016/j.jclepro.2017.03.006

Price, L., & Kendall, A. (2012). Wind Power as a Case Study: Improstadlving Life Cycle Assessment Reporting to Better Enable Meta-analyses. *Journal of Industrial Ecology*, *16*. https://doi.org/10.1111/j.1530-9290.2011.00458.x

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., … Foley, J. A. (2009). A safe operating space for humanity. *Nature*, *461*(7263), 472–475. https://doi.org/10.1038/461472a

Rowley, H. V., Lundie, S., & Peters, G. M. (2009). A hybrid life cycle assessment model for comparison with conventional methodologies in Australia. *International Journal of Life Cycle Assessment*, *14*(6), 508–516. https://doi.org/1007/ s11367-009-0093-5

Smith, H. (1963). Cumulative energy requirements for the production of chemical intermediates and products. In *World Economic Conference*.

Södersten, C. J. (2020). Capital use matrices [exiov3.6_2011_pxp format] retrieved from Zenodo. https://doi.org/10.5281/zenodo.3874309

Södersten, C. J. H., Wood, R., & Hertwich, E. G. (2018). Endogenizing Capital in MRIO Models: The Implications for Consumption-Based Accounting. *Environmental Science and Technology*, *52*(22), 13250–13259. https://doi.org/10.1021/acs.est.8b02791

Södersten, C. J., Wood, R., & Hertwich, E. G. (2017). Environmental Impacts of Capital Formation. *Journal of Industrial Ecology*, *00*(0), 1–13. https://doi.org/10.1111/jiec.12532

Södersten, C. J., Wood, R., & Wiedmann, T. (2020). The capital load of global material footprints. *Resources, Conservation and Recycling*, *158*(March), 104811. https://doi.org/10.1016/j.resconrec.2020.104811

Sonnemann, G., Gemechu Eskinder, D., Sala, S., & Schau, E. M. (2017). Life Cycle Thinking and the use of LCA in policies around the world. In *Life Cycle Assessment: Theory and Practice* (pp. 1–1216). https://doi.org/10.1007/978-3-319-56475-3

Stadler, K., Steen-Olsen, K., & Wood, R. (2014). The “Rest of the world” - estimating the economic structure of missing regions in global multi-regional input-output tables. *Economic Systems Research*, *26*(3), 303–326.

Stadler, K., Wood, R., Bulavskaya, T., Södersten, C. J., Simas, M., Schmidt, S., … Tukker, A. (2018). EXIOBASE 3: Developing a Time Series of Detailed Environmentally Extended Multi-Regional Input-Output Tables. *Journal of Industrial Ecology*, *22*(3), 502–515. https://doi.org/10.1111/jiec.12715

Steen-Olsen, K., Owen, A., Hertwich, E. G., & Lenzen, M. (2014). EFFECTS OF SECTOR AGGREGATION ON CO $_2$ MULTIPLIERS IN MULTIREGIONAL INPUT–OUTPUT ANALYSES. *Economic Systems Research*, *26*(3), 284–302. https://doi.org/10.1080/09535314.2014.934325

Stephan, A., Crawford, R. H., & Bontinck, P.-A. (2018). A model for streamlining and automating path exchange hybrid life cycle assessment. *The International Journal of Life Cycle Assessment*, *24*(2), 237–252. https://doi.org/10.1007/s11367-018-1521-1

Strømman, A. H. (2009). Dealing with double-counting in tiered hybrid life-cycle inventories: a few comments - response. *Journal of Cleaner Production*, *17*(17), 1607–1609. https://doi.org/10.1016/j.jclepro.2009.06.007

Strømman, A. H., Peters, G. P., & Hertwich, E. G. (2009). Approaches to correct for double counting in tiered hybrid life cycle inventories. *Journal of Cleaner Production*, *17*(2), 248–254. https://doi.org/10.1016/j.jclepro.2008.05.003

Strømman, A. H., & Solli, C. (2008). Applying Leontief's price model to estimate missing elements in hybrid life cycle inventories. *Journal of Industrial Ecology*, *12*(1), 26–33. https://doi.org/10.1111/j.1530-9290.2008.00011.x

Strømman, A. H., Solli, C., & Hertwich, E. G. (2006). Hybrid life-cycle assessment of natural gas based fuel chains for transportation. *Environmental Science and Technology*, *40*(8), 2797–2804. https://doi.org/10.1021/es0511523

Suh, S. (2004). Functions, commodities and environmental impacts in an ecological–economic model. *Ecological Economics*, *59*(1), 1–6. https://doi.org/10.1016/j.ecolecon.2003.10.013

Suh, S. (2009). Developing the Sectoral Environmental Database for Input-Output Analysis: Comprehensive Environmental Data Archive of the U.S. In *Handbook of Input-Output Economics in Industrial Ecology* (pp. 689–712).

Suh, S., & Huppes, G. (2005). Methods for life cycle inventory of a product. *Journal of Cleaner Production*, *13*(7), 687–697. https://doi.org/10.1016/j.jclepro.2003.04.001

Suh, S., Lenzen, M., Treloar, G. J., Hondo, H., Horvath, A., Huppes, G., … Norris, G. (2004). System Boundary Selection in Life-Cycle Inventories Using Hybrid Approaches. *Environmental Science and Technology*, *38*(3), 657–664. https://doi.org/10.1021/es0263745

Suh, S., & Lippiatt, B. C. (2012). Framework for hybrid life cycle inventory databases : a case study on the Building for Environmental and Economic Sustainability (BEES) database, 604–612. https://doi.org/10.1007/s11367-012-0393-z

Szyrmer, J. M. (1992). Input—Output Coefficients and Multipliers from a Total-Flow Perspective. *Environment and Planning A*, *24*, 921–937.

Teh, S. H., Wiedmann, T., & Moore, S. (2018). Mixed - unit hybrid life cycle assessment applied to the recycling of construction materials. *Journal of Economic Structures*, *7*(1). https://doi.org/10.1186/s40008-018-0112-4

The Vienna Institute for International Economic Studies. (2017). EU KLEMS. Retrieved from http://www.euklems.net/

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & de Vries, G. J. (2015). An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production. *Review of International Economics*, *23*(3), 575–605.

Tukker, A., & Dietzenbacher, E. (2013). Global multiregional input output frameworks: an introduction and outlook. *Economic Systems Research*, *25*, 1–19. https://doi.org/10.1080/09535314.2012.761179

UNEP/SETAC Life Cycle Initiative. (2012). *Greening the economy through life cycle thinking*.

UNEP. (2003). *Life Cycle Assessment: What it is and How to do it*. Retrieved from http://www.sciencenetwork.com/lca/unep_guide_to_lca.pdf

Vogstad, K.-O. (2009). Input-Output Analysis and LinearProgramming. In *Handbook of Input-Output Economics in Industrial Ecology* (Vol. 23, pp. 801–818). https://doi.org/10.1111/j.1744-7976.1977.tb02886.x

Ward, H., Wenz, L., Steckel, J. C., & Minx, J. C. (2017). Truncation Error Estimates in Process Life Cycle Assessment Using Input-Output Analysis. *Journal of Industrial Ecology*, *22*(5), 1080–1091. https://doi.org/10.1111/jiec.12655

Weidema, B. (2011). Stepping Stones From Life Cycle Assessment to Adjacent Assessment Techniques. *Journal of Industrial Ecology*, *15*(5), 641–642. https://doi.org/10.1111/j.1530-9290.2011.00385.x

Weidema, B., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., … Wernet, G. (2013). *Data quality guideline for the ecoinvent database version 3*. Retrieved from http://www.ecoinvent.org/database/methodology-of-ecoinvent-3/methodology-of-ecoinvent-3.html

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

Wiedmann, T. O., Suh, S., Feng, K., Lenzen, M., Acquaye, A., Scott, K., & Barrett, J. R. (2011). Application of hybrid life cycle approaches to emerging energy technologies - The case of wind power in the UK. *Environmental Science and Technology*, *45*(13), 5900–5907. https://doi.org/10.1021/es2007287

Williams, E. D. (2004). Energy Intensity of Computer Manufacturing: Hybrid Assessment Combining Process and Economic Input-Output Methods. *Environmental Science & Technology*, *38*(22), 6166–6174.

Wing, I. S. (2004). *Computable General Equilibrium Models and Their Use in Economy-Wide Policy Analysis*. *MIT Joint Program on the Science and Policy of Global Change*.

Wolfram, P. (2015). *Carbon Footprint Scenarios for Renewable Electricity Generation in Australia*. Berlin Institute of Technology.

Wood, R., Stadler, K., Bulavskaya, T., Lutter, S., Giljum, S., de Koning, A., … Tukker, A. (2015). Global sustainability accounting-developing EXIOBASE for multi-regional footprint analysis. *Sustainability (Switzerland)*, *7*(1), 138–163. https://doi.org/10.3390/su7010138

WORLD KLEMS Initiative. (2017). WORLD KLEMS. Retrieved November 7, 2019, from http://www.worldklems.net/

Yang, Y., Heijungs, R., & Brandão, M. (2017). Hybrid life cycle assessment (LCA) does not necessarily yield more accurate results than process-based LCA. *Journal of Cleaner Production*, *150*, 237–242. https://doi.org/10.1016/j.jclepro.2017.03.006

Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M., & Meyer, D. E. (2017a). USEEIO: A new and transparent United States environmentally-extended input-output model. *Journal of Cleaner Production*, *158*, 308–318. https://doi.org/10.1016/j.jclepro.2017.04.150

Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M., & Meyer, D. E. (2017b). USEEIOv1.1 [product-by-product industry-technology-assumption coefficients, for year 2017]. Retrived from the US EPA website. Retrieved from https://catalog.data.gov/dataset/useeio-v1-1-matrices

Yu, M., & Wiedmann, T. (2018). Implementing hybrid LCA routines in an input – output virtual laboratory. *Journal of Economic Structures*, *7*(33). https://doi.org/10.1186/s40008-018-0131-1

Zhai, P., Williams, E. D., State, A., Zhai, P., & Williams, E. D. (2010). Dynamic hybrid life cycle assessment of energy and carbon of multicrystalline silicon photovoltaic systems. *Environmental Science & Technology*, *44*(20), 7950–7955. https://doi.org/10.1021/es1026695

ANNEXE A INFORMATIONS SUPPLEMENTAIRES DES ARTICLES

Les informations supplémentaires (supporting information) des trois articles étant trop imposants, elles ont été stockées sur Zenodo : https://zenodo.org/record/3937505

ANNEXE B DEMONSTRATION DU BIAIS D'AGREGATION DE L'IO

CONSTRUCTION D'UN EXEMPLE

Cette annexe a pour but de démontrer le biais d'agrégation de l'IO. Pour ce faire, l'annexe s'appuie sur un exemple. On considère deux technologies : « véhicules à combustion » et « véhicules électriques », lesquels vont être agrégés en un secteur « véhicules ». Les captures d'écran ci-dessous (obtenues avec Jupyter notebook) montrent les caractéristiques des différents véhicules. On considère des intrants d'électricité, d'acier, de moteur à combustion, de batteries lithium-ion et d'autres intrants non spécifiés. La quantité de chaque intrant est générée de façon aléatoire, à part les quantités de batteries lithium-ion et de moteurs à combustion qui sont fixes à zéro pour les véhicules à combustion et véhicules électriques respectivement. Pour chaque fonction de production, une valeur ajoutée est nécessaire pour s'assurer que la somme des intrants et de cette valeur ajoutée soit égale à 1 en tout temps. Cette valeur ajoutée n'a aucune incidence sur la démonstration et n'est donc pas représentée.

```
combustion_vehicles = pd.DataFrame(np.random.uniform(low=0.01, high=0.2, size=(5,)),
                                   columns=['technological coefficients (€)'],
                                   index=['electricity', 'Li battery','combustion engine','steel','other'])
combustion_vehicles.loc['Li battery'] = 0
combustion_vehicles
```

	technological coefficients (€)
electricity	0.107093
Li battery	0.000000
combustion engine	0.132794
steel	0.092755
other	0.110014

```
electric_vehicles = pd.DataFrame(np.random.uniform(low=0.01, high=0.2, size=(5,)),
                                 columns=['technological coefficients (€)'],
                                 index=['electricity', 'Li battery','combustion engine','steel','other'])
electric_vehicles.loc['combustion engine'] = 0
electric_vehicles
```

	technological coefficients (€)
electricity	0.129685
Li battery	0.163865
combustion engine	0.000000
steel	0.167433
other	0.081488

Si l'on considère (à travers une récolte de données par exemple) que le ratio de ventes entre véhicules à combustion et véhicules électriques est de 8 : 2, on obtient la fonction de production suivante pour le secteur agrégé « véhicules ».

```
vehicles = combustion_vehicles*0.8+electric_vehicles*0.2
vehicles
```

	technological coefficients (€)
electricity	0.111611
Li battery	0.032773
combustion engine	0.106235
steel	0.107691
other	0.104309

En définissant des impacts normalisés (prenons le réchauffement climatique pour cet exemple) pour chacun des intrants de notre économie (électricité, batteries, etc.) :

```
normalized_impacts = pd.DataFrame(np.random.rand(5,1), columns=['normalized_impacts (kgCO2e)'],
                                  index=['electricity','Li battery','combustion engine','steel','other'])
normalized_impacts
```

	normalized_impacts (kgCO2e)
electricity	0.294837
Li battery	0.270858
combustion engine	0.923216
steel	0.850489
other	0.206939

on peut calculer les impacts normalisés respectifs des différents véhicules :

```
combustion_vehicles_impacts = normalized_impacts.T.dot(combustion_vehicles)
combustion_vehicles_impacts
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	0.255826

```
electric_vehicles_impact = normalized_impacts.T.dot(electric_vehicles)
electric_vehicles_impact
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	0.241883

et s'assurer que ces impacts correspondent bien à l'impact normalisé du secteur agrégé « véhicules » :

```
vehicles_impacts = normalized_impacts.T.dot(vehicles)
vehicles_impacts
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	0.253038

```
vehicles_impacts == combustion_vehicles_impacts*0.8 + electric_vehicles_impacts*0.2
```

	technological coefficients (€)
normalized impacts (kgCO2e)	True

Nous avons donc créé le secteur agrégé « véhicules » avec succès, en supposant un ratio 8 : 2 entre véhicules à combustion et véhicules électriques.

Regardons maintenant ce qui se passe lorsqu'une demande finale est utilisée. Disons que notre demande finale (consommateurs, gouvernement, etc.) a commandé 8000000€ de véhicules à combustion et 2000000€ de véhicules électriques pour un total de 10000000€ de véhicules. On a donc un impact donné par le secteur agrégé « véhicules » de :

```
aggregated_impact = vehicles_impacts*10000000
aggregated_impact
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.530377e+06

L'impact non agrégé quant à lui donne :

```
final_demand = 8000000*combustion_vehicles+2000000*electric_vehicles
non_agregated_impact = normalized_impacts.T.dot(final_demand)
non_agregated_impact
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.530377e+06

Puisque notre demande finale respectait le ratio 8 : 2 avec lequel le secteur agrégé a été construit, les résultats sont identiques.

ANALYSES RÉTROSPECTIVES ET BIAIS D'AGRÉGATION

Supposons maintenant deux secteurs : transport de marchandises et transport de passagers, tous deux faisant appel aux services fournis par notre secteur agrégé « véhicules » tel que la somme des véhicules à combustion et électriques totale soit dans un ratio 8 : 2. Considérons donc un secteur de transports de marchandises qui, avec les données non agrégées, serait constitué de 4000000€ de véhicules à combustion uniquement et un secteur de transport de passagers qui serait constitué de 4000000€ de véhicules à combustion et de 2000000€ de véhicules électriques. La demande totale sur le secteur agrégé « véhicules » est donc de 8000000€ de véhicules à combustion et 2000000€ de véhicules électriques, autrement dit le ratio 8 : 2 est bien respecté. Et pourtant il y a tout de même un bien d'agrégation, car, pris séparément, les deux secteurs de transport ne respectent pas le ratio 8 : 2. Les quatre captures d'écran ci-dessous montrent la différence entre les impacts agrégés et non agrégés pour ces secteurs.

```
good_transportation = 4000000*combustion_vehicles
good_transportation_impacts = normalized_impacts.T.dot(good_transportation)
good_transportation_impacts
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	1.023305e+06

```
aggregated_impact_goods = normalized_impacts.T.dot(4000000*vehicles)
aggregated_impact_goods
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	1.012151e+06

```
passenger_transportation = 4000000*combustion_vehicles+2000000*electric_vehicles
passenger_transportation_impacts = normalized_impacts.T.dot(passenger_transportation)
passenger_transportation_impacts
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	1.507071e+06

```
aggregated_impact_passenger = normalized_impacts.T.dot(6000000*vehicles)
aggregated_impact_passenger
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	1.518226e+06

Il faut noter néanmoins, que l'impact total lui, ne souffre pas de biais d'agrégation :

```
good_transportation_impacts+passenger_transportation_impacts
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.530377e+06

```
aggregated_impact_goods+aggregated_impact_passenger
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.530377e+06

En quoi est-ce un problème alors ? Dans une étude rétrospective, le but de l'étude est typiquement de regarder les contributeurs majeurs d'un impact, i.e., quels secteurs économiques sont le plus responsable. En ACV, on parlerait d'une analyse de contribution. Or, le biais d'agrégation vient modifier les vraies valeurs de cette analyse de contribution. En suivant notre exemple, l'analyse de contribution effectuée avec les valeurs agrégées nous donne le résultat suivant :

```
aggregated_contribution_good = (aggregated_impact_goods/(aggregated_impact_goods+aggregated_impact_passenger)*100).iloc
aggregated_contribution_passenger = (aggregated_impact_passenger/(aggregated_impact_goods+aggregated_impact_passenger)*
aggregated_contributions = pd.DataFrame([aggregated_contribution_good, aggregated_contribution_passenger],
                                        columns=['contribution (%)'],
                                        index=['goods','passengers'])
aggregated_contributions
```

	contribution (%)
goods	40.0
passengers	60.0

alors que la contribution réelle est :

```
contribution_good = (good_transportation_impacts/(good_transportation_impacts+passenger_transportation_impacts)*100).il
contribution_passenger = (passenger_transportation_impacts/(good_transportation_impacts+passenger_transportation_impact
contributions = pd.DataFrame([contribution_good, contribution_passenger],
                                        columns=['contribution (%)'],
                                        index=['goods','passengers'])
contributions
```

	contribution (%)
goods	40.440825
passengers	59.559175

Le biais d'agrégation de l'IO vient donc fausser les contributions des différents secteurs dans une étude rétrospective, si le ratio appliqué lors de l'agrégation n'est pas respecté *dans la demande de chaque secteur*.

ANALYSES PROSPECTIVES ET BIAIS D'AGRÉGATION

Les analyses prospectives ont typiquement comme but de changer la demande finale pour refléter un changement de consommation par exemple. Ce changement, s'il ne respecte pas le ratio utilisé lors de l'agrégation, faussera les impacts obtenus par une analyse prospective. Supposons donc une demande finale de 9000000€ de véhicules à combustion et 1000000€ de véhicules électriques, soit un ratio de 9 : 1. L'impact agrégé reste identique, car la demande totale est toujours égale à 10000000€.

```
aggregated_impact = vehicles_impacts*10000000
aggregated_impact
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.530377e+06

L'impact non agrégé en revanche a changé :

```
final_demand = 9000000*combustion_vehicles+1000000*electric_vehicles
non_agregated_impact = normalized_impacts.T.dot(final_demand)
non_agregated_impact
```

	technological coefficients (€)
normalized_impacts (kgCO2e)	2.544320e+06

Les résultats obtenus entre les versions agrégée et non agrégée sont différents dus au biais d'agrégation de l'IO.