



HAL
open science

Dépendances technologiques : Développement d'un système de liaison entre le modèle d'Analyse de Cycle de Vie (ACV) et la modélisation MATER

Baptiste de Goër

► **To cite this version:**

Baptiste de Goër. Dépendances technologiques : Développement d'un système de liaison entre le modèle d'Analyse de Cycle de Vie (ACV) et la modélisation MATER. Modélisation et simulation. 2022. hal-03736623

HAL Id: hal-03736623

<https://inria.hal.science/hal-03736623v1>

Submitted on 22 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

GRENOBLE INP – ENSIMAG
ÉCOLE NATIONALE SUPÉRIEURE D'INFORMATIQUE ET DE
MATHÉMATIQUES APPLIQUÉES

Rapport de projet de fin d'études

Effectué au sein de l'équipe de recherche STEEP à l'INRIA
Soutenabilité, Transition, Environnement, Économie biophysique et Politiques locales

Hébergé par le Laboratoire d'Informatique de Grenoble - LIG

En collaboration avec l'Institut des Sciences de la Terre - ISTerre

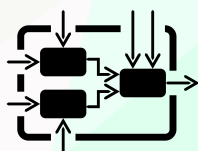
Dépendances technologiques :
Développement d'un système de liaison entre
le modèle d'Analyse de Cycle de Vie (ACV) et
la modélisation MATER

de GOËR Baptiste

3^{ème} année – Option ISI – Semestre PISTE

31 JANVIER 2022 - 29 JUILLET 2022

SOUTENU LE 24 JUIN 2022



STEEP

INRIA

655 Av. de l'Europe

38330 Montbonnot-Saint-Martin

<https://steep.inria.fr/>

Responsables de stage

MANDIL GUILLAUME

TOURANCHEAU BERNARD

Tuteur de l'école

BOUVERET SYLVAIN

Table des matières

Liste des figures	3
Liste des tableaux	3
Coordonnées	5
Glossaire	5
1 Mise en contexte	6
1.1 Contexte socio-écologique	6
1.2 Environnement du stage	6
1.3 Objectifs du stage	6
2 Le projet et ses objectifs	7
2.1 Introduction à l'analyse entrée-sortie	7
2.2 Analyse de Cycle de Vie	8
2.2.1 Objectif	8
2.2.2 Méthode	8
2.2.3 Formalisme mathématique et structure technique	11
2.3 Modélisation MATER	13
2.3.1 Description technique	13
2.4 Objectifs détaillés du projet	14
2.4.1 Lien avec le travail de Jeanne BARRIOL	14
3 Solution technique au projet	15
3.1 Etat de l'art	15
3.2 Outils utilisés	15
3.2.1 Brightway2	15
3.2.2 Docker	17
3.3 Structure logicielle	17
3.3.1 Modèle-Vue-Controller	17
3.3.2 Déploiement	17
3.4 Critères d'évaluation du fonctionnement	18
4 Implémentation du projet	19
4.1 Détail de l'objectif principal du projet	19
4.2 Processus d'implémentation par essai-erreur	21
4.2.1 À l'ordre 1	21
4.2.2 À l'ordre 2	21
4.2.3 À l'ordre n	22
4.2.4 Les cycles	23
4.2.5 Les entrées négatives	25
4.3 Preuve de fonctionnement	25
4.3.1 Invariant	25
4.3.2 Approche de la démonstration	26
5 Tests	27

5.1	Système de test	27
5.1.1	Test unitaire	27
5.1.2	Jeux de test	27
5.2	Problématique technique	27
5.2.1	Explication	28
5.2.2	Approches de résolution	29
5.3	Choix finaux d'implémentation	33
5.3.1	Vérification façon Monte Carlo	33
6	Bilan et suite du projet	35
6.1	Bilan technique du projet	35
6.2	Limites	35
6.3	Où aller pour la suite?	36
6.3.1	Déploiement	36
6.3.2	Interface directe avec le code de MATER	36
6.3.3	Groupement semi-automatique	36
6.3.4	Identification des priorités d'intégration dans MATER	36
6.3.5	Données prédictives	37
7	Évaluation de l'impact socioenvironnemental du projet	38
7.1	Bilan personnel	38
7.2	Structure d'accueil	40
7.2.1	INRIA	40
7.2.2	STEEP	42
7.3	Impact du projet	43
	Références bibliographiques	48
 Annexes		
A	Démonstration formelle de l'algorithme de groupement	51
a	Définition des termes	51
b	Démonstration	52
B	Tableau des indicateurs de test	54
C	Comparatif de la répartition du premier chiffre significatif sur des vecteurs d'erreurs absolues et des vecteurs d'erreurs relatives	55
D	Organisation du travail	56
a	Diagramme de Gantt	56
b	Fonctionnement du projet	56
c	Journée type	57
E	Retour d'expérience sensible du pré-rapport	58

Table des figures

1	Activité entrée-sortie	9
2	Exemple de graphe d'activités	10
3	Schéma structurel de Brightway2	16
4	Exemple de groupement	19
5	Exemple d'activité groupée	20

Liste des tableaux

1	Coordonnées des acteurs du stage	5
2	Glossaire	5
3	Tableau explicatif des entrées et sorties de technosphère et de biosphère	9
4	Liste des impacts et de leur méthode dans le cadre de l'ACV du stage	39
5	Quelques exemples de conséquences des résultats d'une modélisation prospective de transition énergétique	44

Résumé

Aujourd'hui, la principale méthodologie utilisée pour analyser les impacts environnementaux d'un système donné est l'Analyse de Cycle de Vie (ACV). Cette méthodologie multi-étapes et multi-critères est maintenant installée dans le consensus comme fiable dans un objectif de prospective. Dans ce cadre, d'immenses bases de données ont été mises en place. C'est le cas notamment d'Écoinvent, une base de données d'ACV suisse particulièrement complète.

Il y a actuellement un enjeu d'association de ces bases de données, et de cette méthodologie, avec d'autres modèles prospectifs, comme par exemple le modèle MATER proposé par une équipe de recherche du laboratoire ISTerre.

L'objet du stage décrit dans ce document est de développer un programme permettant de grouper des chaînes de production présentes dans Écoinvent, afin que ces groupement correspondent à la structure du modèle MATER, dans l'objectif d'en extraire des données d'ACV et les réutiliser dans ce nouveau cadre.

Le coeur du rapport se concentre sur deux éléments : le développement d'un algorithme fiable de groupement des données, et le système de test mis en place pour assurer cette fiabilité.

Mots clés

Analyse de Cycle de Vie
Multi-regional Assessment of Technologies, Energy and Ressources
Projet de Fin d'Études

Coordonnées

Afin de pouvoir facilement contacter les principaux acteurs du stage, cette section a pour but d'indiquer leurs coordonnées. Le stage est co-encadré par deux tuteurs de stage au niveau de la structure d'accueil STEEP, et d'un encadrant au niveau de l'école.

<i>Nom et prénom</i>	<i>Fonction</i>	<i>Email</i>
Baptiste de GOËR	Stagiaire	baptiste.de-goer-de-herve@grenoble-inp.org
Sylvain BOUVERET	Tuteur d'école	sylvain.bouveret@imag.fr
Guillaume MANDIL	Responsable du stage	guillaume.mandil@univ-grenoble-alpes.fr
Bernard TOURANCHEAU	Responsable du stage	bernard.tourancheau@univ-grenoble-alpes.fr

TABLE 1 – Coordonnées des acteurs du stage

Glossaire

Un certain nombre des termes techniques et des acronymes de ce rapport sont référencés dans cette section.

- STEEP** : *Soutenabilité, Transition, Environnement, Économie biophysique et Politiques locales* ; équipe de recherche interdisciplinaire sur les thématiques d'alternatives socio-techniques et de risques systémiques globaux. Plus d'informations sur <https://steep.inria.fr>
- INRIA** : *Institut National de Recherche en Informatique et en Automatique* ; institut de recherche hébergeant, entre autres, l'équipe STEEP à Grenoble. Plus d'informations sur <https://www.inria.fr>
- LIG** : *Laboratoire d'Informatique de Grenoble* ; laboratoire de recherche hébergeant officiellement le stage. Plus d'informations sur <https://www.liglab.fr>
- ISTerre** : *Institut des Sciences de la Terre* ; institut sur l'étude physique et chimique de la planète Terre. Plus d'informations sur <https://www.isterre.fr>
- MATER** : *Multi-regional Assessment of Technologies, Energy and Ressources* ; Modélisation proposée au sein d'ISTerre par l'équipe d'Olivier VIDAL visant à quantifier la soutenabilité géophysique de scénarios de transition énergétique
- ACV** : *Analyse de Cycle de Vie* ou LCA en anglais (*Life Cycle Analysis*) ; méthodologie visant à estimer les impacts environnementaux d'une unité fonctionnelle (généralement un produit). Voir section 2.2 pour plus de détails sur son fonctionnement.
- ICV** : *Inventaire de Cycle de Vie* ; deuxième étape de l'ACV consistant à remonter l'ensemble des dépendances de la chaîne de production.
- BD** : *Base de Données* ; Ensemble de données spécifiques. Dans le cadre de ce stage, il s'agit de données d'ACV, soit d'activités et de liens entre ces activités.

TABLE 2 – Glossaire

1 Mise en contexte

1.1 Contexte socio-écologique

A l'heure d'une crise multidimensionnelle [1] : a minima sociale [2], écologique [3] et politique [4] ; il est urgent de repenser les modèles de sociétés humaines en prenant en compte les limites planétaires [5], mais également les droits sociaux [6], le tout dans un système de prise de décisions participatif [7].

L'une des tendances actuelles en terme d'écologie est de parler de transition énergétique [8][9], ce sujet apparaissant régulièrement dans le débat public, considérée comme une part importante de la solution à la neutralité carbone [10]. Il faut malgré tout garder en tête que cette approche des problèmes environnementaux est très controversée [11].

La sphère de la recherche scientifique est donc en devoir de mettre à contribution ses capacités, notamment de modélisation, afin de gérer cette restructuration de la société et d'aider à la prise de décision de mesures politiques.

1.2 Environnement du stage

L'équipe STEEP (Soutenabilité, Transition, Environnement, Économie biophysique et Politiques locales) est une équipe de recherche pluridisciplinaire ayant pour objectif de développer des outils d'aide à la prise de décision de politiques écologiques co-construites avec les parties-prenantes. Elle s'est formée suite au constat exprimé dans la section 1.1.

Cette équipe se divise en deux axes de recherche. D'une part l'axe dit AST (Alternatives Socio-Techniques) qui cherche à évaluer la soutenabilité biophysique de modèles de société alternatifs, et d'autre part l'axe dit RSG (Risques Systémiques Globaux) qui vise à étudier les risques à court et long terme du système de société courant.

Mon stage se passe au sein la branche AST de l'équipe de recherche.

STEPP travaille également en étroite collaboration avec l'équipe d'Olivier VIDAL à ISTerre, qui développe le projet MATER, une modélisation visant à évaluer la quantité de ressources nécessaires pour la mise en place de scénarios de transition énergétique.

1.3 Objectifs du stage

Le cœur de mon rôle en tant que stagiaire chez STEEP consiste à travailler sur les bases de données d'Analyse de Cycle de Vie (ACV), afin de développer des outils permettant d'associer ces bases de données, riches en informations, avec d'autres types de modélisation. En pratique, j'ai travaillé sur cette association avec la modélisation du projet MATER, en collaboration avec Jeanne BARRIOL, stagiaire au sein d'ISTerre.

2 Le projet et ses objectifs

La section 1.3 indique que le projet a pour but de faire du lien technique entre des bases de données d'ACV et la modélisation MATER. Il convient donc a minima d'expliquer comment fonctionnent ces deux systèmes, avant de présenter le rôle précis du stage.

Les fondements de la compréhension du fonctionnement des bases de données d'ACV et des outils permettant d'en faire un usage non conventionnel sont tirés du travail d'un stagiaire m'ayant précédé à STEEP (et ayant donc défriché le terrain) : Léo GIGANDON [12].

2.1 Introduction à l'analyse entrée-sortie

Le travail d'Alexandre BORTHOMIEU, anciennement stagiaire à STEEP, aujourd'hui doctorant, a servi de base à la rédaction de cette section [13].

Pour bien comprendre la logique de la structure de l'Analyse de Cycle de Vie, il est important de revenir aux fondements de l'analyse entrée-sortie (ou input-output). Cette section fait office d'introduction à la méthode.

Cette logique d'analyse a été fortement propulsée par le travail en économie de Wassily Leontief (ayant reçu le prix d'économie de la Banque de Suède en la mémoire d'Alfred Nobel), grâce à ses travaux du milieu du XX^{ème} siècle [14].

L'objectif de l'analyse entrée-sortie est d'estimer la demande totale en ressources d'une économie réelle. La première étape consiste à construire un tableau entrées-sorties, donnant pour chaque secteur d'activité la quantité de chaque autre secteur d'activité nécessaire à sa production.

Prenons l'exemple que Leontief développe dans son papier pour expliquer sa logique. Il construit une économie très simple avec deux secteurs d'activité : l'agriculture (qui produit du blé) et la manufacture (qui produit du tissu).

Pour produire 100 boisseaux de blé, il faut :

- 25 boisseaux de blé ($\frac{25}{100} = 0.25$ par unité produite)
- 14 mètres de tissu ($\frac{14}{100} = 0.14$ par unité produite)

Pour produire 50 mètres de tissu, il faut :

- 20 boisseaux de blé ($\frac{20}{50} = 0.40$ par unité produite)
- 6 mètres de tissu ($\frac{6}{50} = 0.12$ par unité produite)

On obtient alors la matrice A de dépendances des secteurs par unité de production (avec Ag pour "Agriculture" et Ma pour "Manufacture") :

$$A = \begin{matrix} & \begin{matrix} Ag & Ma \end{matrix} \\ \begin{matrix} Ag \\ Ma \end{matrix} & \begin{bmatrix} 0.25 & 0.40 \\ 0.14 & 0.12 \end{bmatrix} \end{matrix}$$

Les colonnes de cette matrice représentent la quantité de chaque secteur d'activité nécessaire pour produire une unité du secteur concerné.

La matrice de Leontief inverse (notée L) est ensuite calculée comme suit :

$$L = (I - A)^{-1}$$

$$L \simeq \begin{matrix} & Ag & Ma \\ Ag & \begin{bmatrix} 1.457 & 0.662 \end{bmatrix} \\ Ma & \begin{bmatrix} 0.232 & 1.242 \end{bmatrix} \end{matrix}$$

Cette matrice donne, pour chaque colonne, la quantité totale de production nécessaire de chaque secteur pour produire une unité du secteur concerné. Par exemple, au total, il faudra produire 1.457 unités de blé et 0.232 unités de tissu pour obtenir une unité de blé (grossièrement, les 0.457 unités de blé supplémentaires ont servi à nourrir les agriculteurs-rices et les 0.232 unités de tissu ont servi à les habiller).

C'est sur cette logique que fonctionne l'Analyse de Cycle de Vie.

2.2 Analyse de Cycle de Vie

2.2.1 Objectif

L'analyse de cycle de vie est une méthodologie multi-étapes et multi-critères visant à évaluer les impacts environnementaux de ce que l'on appelle une unité fonctionnelle.

L'unité fonctionnelle est définie par un objectif (ou une fonctionnalité) et une durée [15]. Elle permet notamment de faire le comparatif entre deux méthodes ou produits pour atteindre l'objectif (accomplir la fonctionnalité). Prenons un exemple : si l'on désire effectuer l'ACV d'un téléphone portable, alors l'unité fonctionnelle peut être "appeler et envoyer des messages où que l'on soit pendant 3 ans". Il est ensuite possible de comparer les impacts environnementaux de cette unité fonctionnelle si l'on se sert d'une part d'un smartphone, ou d'autre part d'un téléphone à touches pour l'atteindre.

C'est une méthode qui est très utilisée aujourd'hui dans l'industrie pour faire de l'optimisation environnementale de procédés, ou pour atteindre des objectifs d'économie circulaire. Plus précisément, elle est généralement exploitée soit dans une optique préventive (écoconception), soit dans une logique analytique de procédés déjà en place (bilan environnemental).

L'un des points forts de cette méthode, c'est qu'elle permet d'identifier sur quelle étape du cycle de vie d'un produit se concentrent ses impacts environnementaux. C'est ainsi que l'on sait, par exemple, que le cœur des impacts environnementaux du numérique se trouve dans la fabrication des composants électroniques [16].

2.2.2 Méthode

Structure

Afin de pouvoir identifier les impacts environnementaux d'un produit ou d'un service, la méthode ACV se veut la plus exhaustive possible. Elle cherche donc à mesurer toutes les interactions entre l'environnement et le procédé étudié à toutes les étapes de son cycle de vie (méthodologie dite de "cradle-to-grave") [17].

Pour ce faire, est utilisée une structure atomique d'activités, correspondant à une logique entrée-sortie (ie. les secteurs de la section 2.1). Une activité récupère des éléments en entrée et renvoie un produit en sortie (voir figure 1).

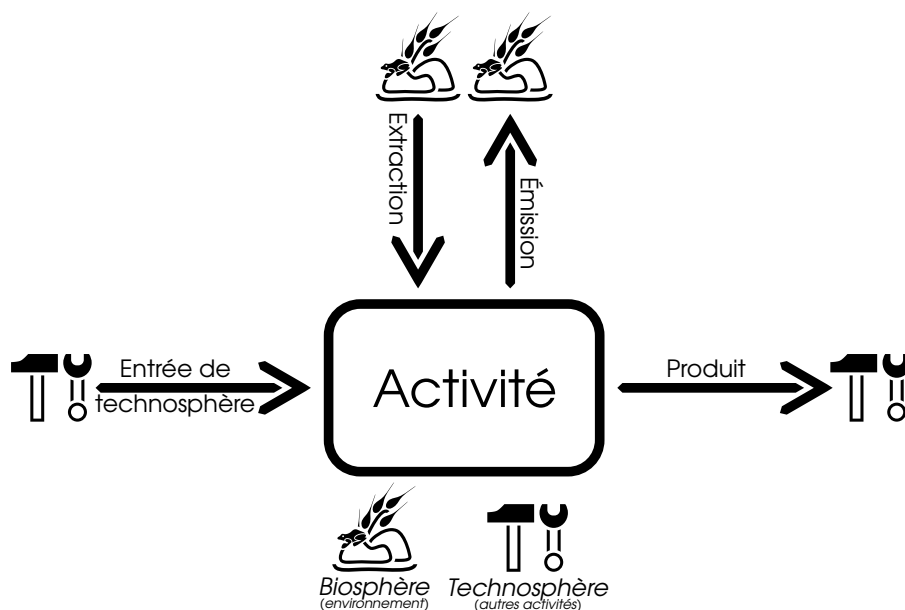


FIGURE 1 – Activité entrée-sortie

Précision sur les activités à sorties multiples

Le formalisme de l'ACV ne permet pas à une activité de posséder plusieurs produits en sortie. Lorsque malgré tout une activité possède plusieurs sorties (notamment dans le cas où elle génère des déchets), un produit est considéré comme étant la sortie de l'activité et les co-produits sont alors considérés comme des entrées venant des filières de traitement. Ces co-produits ont un montant d'entrée négatif et les filières de traitement dont ils sont issus ont un montant de sortie négatif. Cette logique de double négatif se veut consistante avec le formalisme de Leontief [18].

Prenons l'exemple d'une activité qui lors de son processus rejette des eaux usées. Elle a besoin d'une autre activité correspondant à une station d'épuration pour les traiter. La station d'épuration est alors mise en entrée de l'activité, avec un montant négatif correspondant à la quantité d'eau usée rejetée. C'est à lire comme "Pour que l'activité fonctionne, elle a besoin d'une certaine quantité de service d'épuration d'eau".

Ces entrées et ces sorties sont classées en 2 catégories distinctes : la technosphère et la biosphère.

	Technosphère	Biosphère
Entrée	Élément issu de la sortie d'une autre activité	Ressource prélevée dans l'environnement
Sortie	Produit destiné à la consommation, soit par une autre activité, soit comme produit fini	Émission dans l'environnement

TABLE 3 – Tableau explicatif des entrées et sorties de technosphère et de biosphère

Précision sur les flux de biosphère

Le formalisme de l'ACV considère toutes les interactions avec l'environnement comme des entrées. Les émissions dans l'environnement sont alors traitées comme des entrées négatives (dans la même logique que pour les co-produits).

Il est possible de visualiser l'ensemble comme un graphe orienté : les sommets non-flots et non-sources sont les activités de technosphère, les arcs sont les flux de technosphère ou de biosphère, les sommets sources sont des interactions avec l'environnement¹ et les sommets flots sont soit des interactions avec l'environnement, soit la demande finale. La figure 2 présente un exemple de cette représentation.

La pondération des arcs correspond à la quantité de l'activité source nécessaire à la production d'une unité de l'activité cible. Par le terme "activité" s'entendent ici également les interactions avec l'environnement. La pondération vers la demande finale correspond à la quantité de l'activité source demandée en sortie. C'est par là que l'on retrouve les matrices entrées-sorties présentées en section 2.1 : une activité a besoin d'une certaine quantité des autres activités² pour fonctionner.

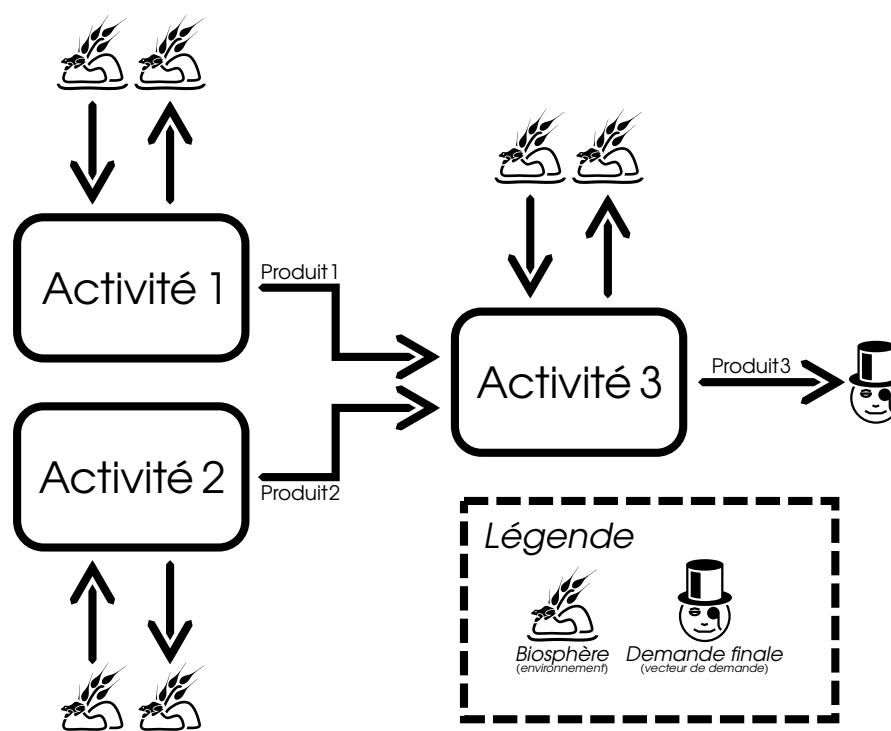


FIGURE 2 – Exemple de graphe d'activités

Inventaire de Cycle de Vie (ICV)

Une fois la base de données avec toutes les activités recensées, indiquant pour chacune ses entrées et sorties, il faut préciser une chaîne de production, de distribution, d'usage et de traitement des déchets (suite d'activités représentant l'ensemble des étapes du cycle de vie du produit ou du service). Cette chaîne correspond à un ensemble de produits de la base de données, associés à une certaine quantité (cela correspond à la *demande finale* de la figure 2). C'est ce que l'on appelle le tableau de demande (voir section 2.2.3).

L'ACV va ensuite remonter tout le graphe d'activités pour chaque élément du tableau de demande, afin de déterminer l'ensemble des flux de biosphère associés et essayer d'inventorier

1. Par "interactions avec l'environnement", s'entend un processus d'émission ou d'extraction, par exemple des émissions de CO_2 , des rejets d'ammoniac, l'extraction de ressources minières, l'usage des sols, etc. Ces processus correspondent au milieu de la "Biosphère" par différenciation ceux du milieu de la "Technosphère" qui correspondent à des activités de transformation humaines.

2. Éventuellement également d'elle-même, dans ce cas il y aura une boucle unitaire dans le graphe.

de la façon la plus exhaustive possible les interactions entre l'environnement et l'ensemble du cycle de vie de l'unité fonctionnelle (dans la figure 2, cela correspond à l'ensemble des flèches en lien avec la biosphère).

C'est le principe du Life Cycle Inventory (LCI), ou inventaire de cycle de vie en français (ICV).

Précision sur l'inventaire

Ce qui est décrit ici correspond à l'inventaire de biosphère. Il existe une phase intermédiaire qui correspond à l'inventaire de technosphère, c'est-à-dire les quantités totales de chaque activité du graphe nécessaire à la production du tableau de demande. Ces quantités équivalent à la matrice de Leontief inverse (voir section 2.1). Le détail se trouve dans la section 2.2.3.

Calcul d'impact

Une fois l'inventaire des flux de biosphère déterminé, il faut transformer ces valeurs en indicateurs d'impacts environnementaux. Typiquement, pour les émissions de gaz à effet de serre, en équivalent CO_2 . Cela se traduit simplement par un ensemble de facteurs de pondération pour chaque flux de biosphère, pour chaque impact. L'ensemble de ces facteurs de pondération liés aux flux de biosphère s'appelle la méthode de calcul d'impact.

L'on obtient ainsi, pour une unité fonctionnelle donnée, caractérisée par un inventaire de cycle de vie, l'ensemble de ses impacts, répartis par catégories d'impact et par placement dans la chaîne de cycle de vie.

Il y a beaucoup à dire sur cette étape fondamentale de l'ACV, mais celle-ci ne concerne pas le stage, inutile donc de la développer davantage.

2.2.3 Formalisme mathématique et structure technique

Formats de base de données

Il existe de nombreuses bases de données d'ACV. Elles peuvent être publiques ou internes à une société, gratuites ou payantes, libres de droits ou propriétaires. Certaines sont spécialisées comme Agribalyse [19], développée par l'ADEME (Agence De l'Environnement et de la Maîtrise de l'Énergie) et destinée au secteur de l'agriculture. D'autres se veulent le plus générique possible comme Ecoinvent [20], qui fait figure de base de référence pour la descriptions de procédés technologiques.

Ces bases de données ne sont pas dynamiques : les formats sont généralement des formats XML ou CSV, qui sont ensuite importés dans les logiciels de traitement d'ACV.

Il n'y a pas de format universel, Ecoinvent utilise par exemple un format XML appelé EcoSpold2. Les logiciels d'ACV ont généralement leur propre format, comme Simapro qui utilise un format CSV appelé Simapro CSV pour les imports et les exports.

Représentation des Inventaires de Cycle de Vie

Les inventaires de cycle de vie peuvent être représentés sous forme matricielle [21].

Soit n le nombre d'activités dans la base de données.

Soit m le nombre de flux de biosphère existants (ou nécessitant d'être calculés) dans la base de données.

Soit A une matrice de taille $n \times n$ possédant à chaque position $A_{ij, i \neq j}$ une valeur négative (ou nulle) représentant la quantité de produit issu de l'activité i nécessaire à la production d'une unité de produit de l'activité j . Chaque colonne correspond à la recette de production de l'activité j . $A_{jj} = 1$. On appelle cette matrice la matrice de technosphère³.

Soit B une matrice de taille $m \times n$ possédant à chaque position B_{ij} une valeur signée représentant la quantité de flux de biosphère i utilisée (ou émise) par l'activité j . On appelle cette matrice la matrice de biosphère.

Soit f un vecteur de taille n représentant la quantité de chaque activité nécessaire pour générer le produit final. On appelle ce vecteur le tableau de demande.

Soit g un vecteur de taille m représentant la quantité totale de chaque flux de biosphère utilisé par le produit final au cours de son cycle de vie.

On a alors l'équation :

$$g = BA^{-1}f$$

Cette technique de calcul d'impact a l'avantage d'être très rapide à la résolution, d'autant plus que les matrices sont creuses, mais a le défaut de ne pas permettre de désagréger l'origine des impacts. Dans les faits, l'un des avantages d'effectuer des ACV est de connaître ladite origine. Ainsi, un-e utilisateur-riche de logiciel d'ACV indique manuellement quelle partie du tableau de demande correspond à quelle partie du cycle de vie du produit ou du service, ce qui permet d'effectuer des calculs séparés. La chaîne d'impact peut également généralement être remontée dynamiquement, au moyen de seuils de coupure d'impact et/ou d'une certaine profondeur de récursion.

Pondération linéaire du calcul d'impact

Les calculs d'impact sont linéaires :

Soit i une catégorie d'impact.

Soit c^i un vecteur de taille m comprenant à la position j le coefficient d'impact du flux de biosphère j pour la catégorie d'impacts i .

Soit h^i la valeur de la catégorie d'impact i .

On a alors :

$$h^i = \sum_{j=0}^{m-1} c_j^i g_j$$

Apparté sur les limites de cette méthode

Le fait que les calculs d'impact soient une simple multiplication de coefficients en fin de remontée de la chaîne est clairement une limite à la méthodologie classique d'ACV [22]. En effet, par exemple, un rejet d'ammoniac ne va pas avoir le même impact en fonction du type de sol dans lequel il est émis. La façon de limiter ce problème qui est aujourd'hui utilisée est de définir des flux de biosphère différents en fonction du milieu d'émission et ainsi de les pondérer différemment. Mais poussé à l'extrême, ce raisonnement voudrait que chaque flux de biosphère de chaque activité ait son propre coefficient, ce qui n'est en pratique pas implémentable, car il faudrait une quantité astronomique de coefficients, dépendant de chaque base de donnée et de chaque méthode de calcul d'impact.

3. Il faut se méfier, car cette matrice de technosphère ne correspond pas exactement à la matrice A de la méthode de Leontief, mais à la matrice générée par $I - A$ dans son approche (voir section 2.1). Le choix de nommer dans les deux cas la matrice A est confusant, mais il semblerait qu'il s'agisse de la notation standard.

2.3 Modélisation MATER

Le projet MATER est un modèle dynamique lancé par Olivier VIDAL de l'Institut des Sciences de la Terre (ISTerre) visant à estimer les besoins en ressources d'une transition énergétique.

La description qui suit est librement inspirée de la fiche de stage de Jeanne BARRIOL [23].

Le projet de modélisation MATER (aussi appelé DyMEMDS) a un quadruple objectif prédictif vis-à-vis d'un changement d'infrastructure énergétique :

1. Estimer les quantités de matières premières et recyclées nécessaires en fonction de l'évolution des réserves,
2. Estimer les émissions de CO_2 locales et importées liées à la transition,
3. Estimer la quantité d'énergie nécessaire à la mise en place de la nouvelle infrastructure,
4. Estimer les impacts d'une telle transition sur le réchauffement global et la quantité d'eau.

2.3.1 Description technique

La partie technique de la description du modèle MATER est issue en très grande partie du rapport de stage de François VERZIER [24]. Le fonctionnement de la modélisation n'étant aujourd'hui pas encore entièrement public, celui-ci ne sera que peu détaillé et les éléments présentés seront centrés essentiellement sur ce qui concerne directement ce stage.

L'une des principales différences entre cette modélisation et un modèle ACV est que le projet MATER possède une dimension d'estimation temporelle de changement des ressources disponibles des technologies. Il s'agit en effet d'un modèle dit stock-flux (bottom-up principalement, mais les manques sont comblés par une approche top-down), qui étudie l'évolution de stocks de variables en fonction d'une variation temporelle associée à un flux. Cette dimension temporelle est complètement absente des modèles d'ACV, purement quantitatifs et non prédictifs.

Une autre différence est la granularité du modèle. Un modèle ACV se veut le plus désagrégré possible, là où le modèle MATER possède une structure atomique plus large.

Plus spécifiquement, la modélisation MATER s'intéresse à 4 grands secteurs :

- Le transport,
- Le bâtiment,
- L'industrie,
- La production énergétique.

Ces secteurs possèdent en leur sein un certain nombre de *Technical Unit* (TU), par exemple des infrastructures de production énergétique (ex : éoliennes), des technologies de transport (ex : camions), des bâtiments (ex : immeubles), des types d'industries, etc.

Ces TU sont produites à partir de *Raw Materials* (RM), principalement le béton, l'acier, l'aluminium et le cuivre.

Ces RM sont des alliages construits à partir de *Ressources* qui correspondent à des ressources naturelles extraites de l'environnement.

Une partie des TU en fin de vie est recyclée en RM et limite ainsi une part de l'extraction nécessaire à leur renouvellement. C'est ce qui s'appelle la boucle de recyclage.

Là où l'ACV peut posséder des centaines d'étapes, seules 4 sont modélisées par MATER. Par exemple, l'ACV va distinguer l'opération de minage d'un minerai et son raffinage, par contre MATER n'y considère qu'un unique bloc d'extraction de la matière.

2.4 Objectifs détaillés du projet

Dans un contexte de mise en relation de la méthodologie ACV et du projet MATER, suite à de nombreux échanges avec les commanditaires, il a été déterminé un objectif clair pour le stage.

Pour cela, nous utiliserons la base de données d'ACV Ecoinvent : il s'agit d'un standard dans le milieu, très complet, notamment quand il est question d'extraction minière.

Il m'est demandé de développer un système ayant un triple objectif :

1. Permettre de grouper séquentiellement les activités d'Ecoinvent, afin d'avoir une structure qui corresponde à celle de MATER. Ce groupement peut se faire de trois manières différentes :
 - Manuellement
 - Automatiquement
 - Semi-automatiquement
2. Proposer une interface avec le code Python utilisé pour la modélisation MATER.
3. Identifier quels éléments de la base de données d'ACV ne sont pas pris en compte dans la modélisation MATER et ont un impact a priori non négligeable sur le résultat final.

Plus précisément, à court terme, ce qui intéresse l'équipe d'ISTerre, c'est de pouvoir récupérer les flux de biosphère directs de groupes d'activités (qui correspondraient à leur production de TU, production de RM, extraction de Ressources et recyclage des TU), et de préférence de les extraire avec des données spécifiques pour chaque année. La logique bottom-up du modèle MATER demande en effet énormément de données.

Ce stage devra avoir a minima développé un outil permettant de grouper les activités d'Ecoinvent pour en extraire les flux de biosphère directs, pour l'année de validité de la base de données d'ACV. Les données pour les autres années pourront à l'avenir soit être extraites de versions différentes d'Ecoinvent, soit être estimées de façon prédictive (voir section 6.3).

2.4.1 Lien avec le travail de Jeanne BARRIOL

Comme évoqué précédemment, ce stage s'effectue main dans la main avec une stagiaire hébergée par ISTerre, Jeanne BARRIOL.

Son rôle, en tout cas la partie qui est en lien avec mon travail, est d'identifier quels groupements faire dans la base de données d'ACV pour que ceux-ci correspondent au mieux à la structure de fonctionnement de MATER, puis d'effectuer lesdits groupements à l'aide de l'outil que je développe pour en extraire les données.

Elle essaye également d'identifier les seuils de coupure et les invariants au sein de la base de donnée d'ACV, ce qui permettrait d'automatiser au mieux les groupements à effectuer.

3 Solution technique au projet

3.1 Etat de l'art

Le travail sur les fondements techniques sur lesquels ce stage s'appuie a été effectué en 2018 par Léo GIANDON [12]. Il a pré-identifié les outils utilisés pour travailler de façon non conventionnelle avec les bases de données d'ACV, notamment **Brightway2** (voir section 3.2.1).

Bien entendu, ce n'est pas la première fois que des chercheurs-ses cherchent à coupler les bases de données d'ACV et d'autres types de modélisation, notamment des modélisations temporelles. On peut, par exemple, citer, dans les travaux les plus récents, le travail sur la technologie *Presmise* [25], qui vise à combiner les informations comprises dans la base de données Ecoinvent avec des scénarios créés par IAM (Integrated Assessment Modeling) [26]. Dans leur cas, l'objectif principal est d'intégrer la base de données d'ACV à une prospective temporelle. C'est une question qui se pose également pour le couplage MATER et ACV, mais que ce stage n'explore pas, faute de temps. Ceci-dit, le papier en question parle d'une technologie, *Wurst* [27], qui pourrait servir à un futur stage sur la thématique (voir section 6.3).

La part technique du stage la plus importante repose sur le travail de Chris MUTEL, qui a développé au cours de ses activités de recherche un module **Python** ayant pour objectif de détourner l'usage classique des bases de données d'ACV [28] (voir section 3.2.1).

3.2 Outils utilisés

3.2.1 Brightway2

Présentation

Brightway2 [29] est un framework **Python** open-source développé par Chris MUTEL destiné au traitement avancé et non-conventionnel des bases de données d'ACV.

Par "non-conventionnel", s'entend ici que **Brightway2** permet d'avoir la main mise sur les structures de données et les méthodes de calcul utilisées dans le développement de logiciels d'ACV, là où un logiciel d'ACV classique permet avant tout d'effectuer des ACV, sans chercher à permettre de détourner leur fonctionnement et leurs données.

Pourquoi Brightway2 ?

Il existe de nombreux logiciels permettant d'exploiter les bases de données d'ACV, avec chacun leur façon de fonctionner [30]. Les logiciels propriétaires les plus connus sont *SimaPro* et *GaBi* [31]. Il existe également une solution open-source puissante, *OpenLCA* [32].

Ces solutions clé-en-main ne nous concernent pas vraiment lors de ce stage, car d'une part, elles n'ont pas pour vocation d'offrir une API (Interface de Programmation d'Application) complète permettant de développer son propre module ; et d'autre part, elles ne cherchent qu'à faire de l'ACV conventionnelle, et non à travailler de façon détournée avec les bases de données.

Il convient donc de d'utiliser une solution non-conventionnelle : **Brightway2**. Cette solution semble être la meilleure pour plusieurs raisons :

1. Il s'agit d'un framework, qui permet donc de développer facilement ses propres outils.

2. Il est développé en Python, or la modélisation MATER également⁴.
3. Il a été structurellement conçu pour détourner les bases de données d'ACV afin de les exploiter de façon non-conventionnelle, il est donc beaucoup plus flexible que les logiciels conventionnels.
4. Il est open-source (et donc, entre autre, relativement pérenne).
5. Il possède une interface graphique : Activity Browser [33], ce qui n'est pas indispensable pour le stage, mais qui permet de visualiser rapidement et facilement si la solution développée fonctionne correctement (qualitativement), et d'aider au debugging le cas échéant.

Fonctionnement technique

Brightway2 fonctionne par projets. Chaque projet comprend une ou plusieurs bases de données d'ACV, des méthodes de calcul d'impact et d'autres objets nécessaires à son fonctionnement (voir figure 3) [34].



FIGURE 3 – Schéma structurel de Brightway2
Librement inspiré d'un schéma de Chris MUTEL.

4. Qu'il soit développé dans ce langage a en plus l'avantage de ne pas me demander de me familiariser avec un langage que je ne maîtrise pas (gain de temps).

Notons que les bases de données avec lesquelles travaille **Brightway2** sont des bases de données dynamiques **SQLite**. En effet, initialement les bases de données d'ACV ne sont pas dynamiques, mais pour travailler avec, **Brightway2** les importe dans son propre format de base de données **SQLite**.

3.2.2 Docker

Présentation

Docker est un système de container **Linux** standard. Il permet entre autre de déployer plusieurs applications sur un même serveur, tout en garantissant une certaine étanchéité entre ces dernières pour des raisons de sécurité. Le système **Docker** est moins gourmand en ressources qu'une machine virtuelle, car il ne repose pas sur un système d'émulation.

Pourquoi Docker ?

Afin de travailler sur les bases de données d'ACV avec **Brightway2**, il était important d'apporter des modifications aux bases de données existantes, et donc de conserver une base de données modifiée dynamiquement sous la main. Il ne serait pas pertinent d'exporter et d'importer la base de données modifiée à chaque fois que quelqu'un-e veut travailler dessus.

À partir de ce constat est venu l'idée d'héberger le projet **Brightway2** concerné sur un serveur, afin de travailler dessus à distance et ainsi avoir, d'une part, une base de donnée commune à tous les acteurs et toutes les actrices du projet, et d'autre part de ne pas avoir à effectuer systématiquement les opérations très lourdes d'import et d'export des bases de données dans le système.

Le déploiement sur le serveur se fait via **Docker**, d'abord parce qu'il s'agit d'un système efficace qui a fait ses preuves, mais également parce que **Brightway2** propose déjà dans sa documentation un **Dockerfile** [29].

3.3 Structure logicielle

3.3.1 Modèle-Vue-Controller

Le cœur de l'application sera développée selon une structure MVC (Modèle-Vue-Controller) [35].

Cette structure a l'avantage d'être modulable. En effet, elle peut permettre notamment de faire évoluer facilement l'interface (dans un premier temps en lignes de commandes, puis éventuellement web dans un second temps), mais également de facilement développer une API qui permettrait de faire communiquer à terme le modèle **MATER** et la base de données modifiée d'ACV.

3.3.2 Déploiement

Comme indiqué précédemment, le déploiement de l'application se fera via **Docker**. Un script d'initialisation a été développé pour permettre notamment l'importation de la base **Ecoinvent** dans le système, puis un script de démarrage de l'application permet de lancer le système. Le tout se déploie via le **Dockerfile**, après avoir indiqué les bons paramètres dans un fichier de configuration.

3.4 Critères d'évaluation du fonctionnement

Le bon fonctionnement du système pourra se vérifier qualitativement via l'interface graphique Activity Browser. Bien entendu, une vérification qualitative est insuffisante bien qu'utile. Des jeux de test unitaires seront donc mis en place. La section 5 est dédié à cette question.

4 Implémentation du projet

4.1 Détail de l'objectif principal du projet

Le cœur de la problématique du projet est le suivant : au sein du graphe pondéré d'activités (comme présenté en section 2.2.2) que représente la base de données Ecoinvent, -près de 15000 activités recensées dans la version 3.4 avec laquelle je travaille-, il faut un algorithme qui permette de grouper un sous-ensemble du graphe.

Ce groupement correspond à une nouvelle activité dans la base de données, reliée au reste du graphe par les liens des activités qui lui sont internes, avec les activités qui lui sont externes.

Un sous-ensemble du graphe à grouper doit cependant répondre à certaines spécificités :

1. Il ne peut pas avoir plusieurs sorties. C'est-à-dire qu'au sein du groupe, une et une seule activité voit sa sortie être considérée comme étant la sortie du groupe. Nous nommerons cette activité l'activité terminale.
2. Quel que soit le sommet du graphe interne au groupe, il doit exister une chaîne menant à l'activité terminale. Entre autres, le graphe doit être connexe, et l'activité terminale est la seule à pouvoir n'être la source d'aucune autre activité interne au groupe.

Ces contraintes sont là pour des raisons de praticité technique, mais ne sont pas contraignantes dans l'usage. En effet, il n'y a aucune réalité physique à des groupements d'activités ayant de telles propriétés.

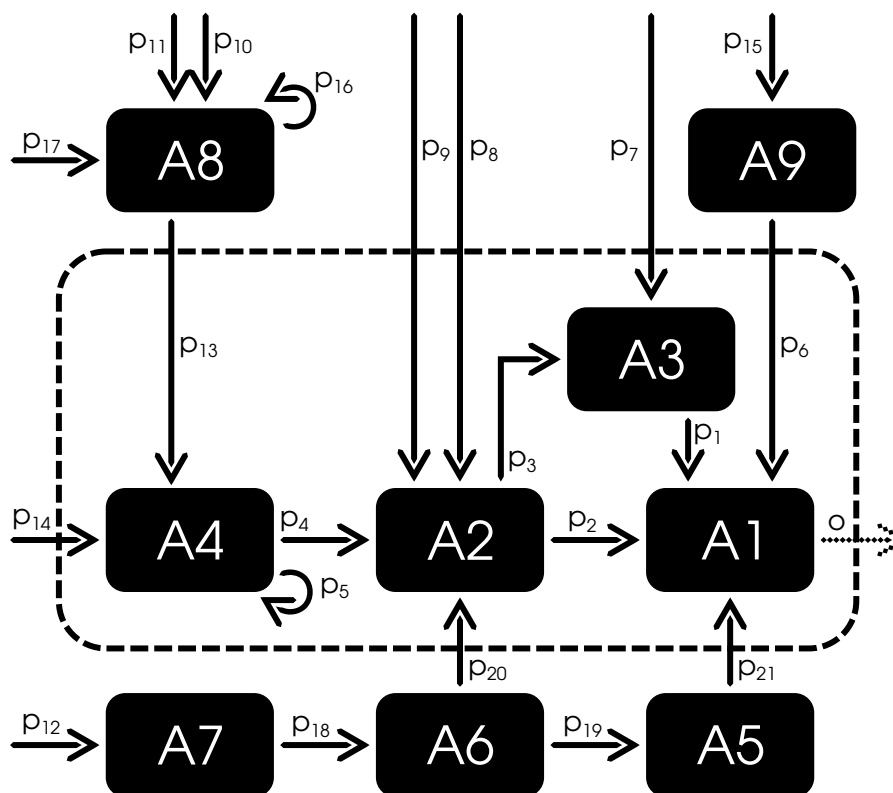


FIGURE 4 – Exemple de groupement

Travaillons sur un exemple : la figure 4.

Dans cet exemple, la base de données possède un ensemble de 9 activités et 21 flux (ou échanges). Les flux depuis l'extérieur peuvent être soit des flux de biosphère, soit des flux reliant

les activités à d'autres activités d'une base de données plus grande, peu importe pour l'exercice qui nous intéresse. Considérons ici qu'il s'agit de flux de biosphère (reliés à l'environnement). Nous sommes donc dans un système fermé.

Chaque échange entre activités est caractérisé par son activité de sortie, son activité d'entrée (qui peut être l'environnement) et un montant, lequel correspond à l'activité d'entrée nécessaire à la production d'une unité de l'activité de sortie. Ces montants sont notés p_i pour chaque flux $i \in \{1, \dots, 21\}$.

Point d'attention sur la sémantique des flèches

Dans ce type de représentation graphique, on trouve deux types de flèches n'ayant pas la même signification. Les arcs pleins correspondent à ce qu'on appelle un échange (ou flux). Ils représentent une certaine quantité de l'activité source nécessaire à la production d'une unité de l'activité cible. Le produit d'une même activité peut donc être nécessaire en différentes quantités à différentes activités. Par exemple, une activité qui produit des batteries pourra être nécessaire en entrée d'une activité produisant des voitures électriques et en entrée d'une activité produisant des smartphones.

En revanche, une activité donnée ne produit bien qu'un seul type de produit (dans l'exemple ci-dessus des batteries), comme indiqué dans un aparté de la section 2.2.2. Cette sortie unitaire, non reliée directement à d'autres activités, est représentée dans le schéma par la flèche en pointillés.

Au sein de cette base de données, l'objectif est de grouper un sous-ensemble formé par les activités $A1$, $A2$, $A3$ et $A4$. La sortie de ce groupement correspond à la sortie d' $A1$: o .

Pour ce groupement, il y a 8 flux qui vont de l'extérieur du groupement vers son intérieur : les flux 6, 7, 8, 9, 13, 14, 20, 21 avec leur montant associé : $p_6, p_7, p_8, p_9, p_{13}, p_{14}, p_{20}$ et p_{21} .

L'activité groupée, que l'on nommera Ag , aura donc 8 flux en entrée, dont nous nommerons les montants $p_6^g, p_7^g, p_8^g, p_9^g, p_{13}^g, p_{14}^g, p_{20}^g$ et p_{21}^g . Voir figure 5.

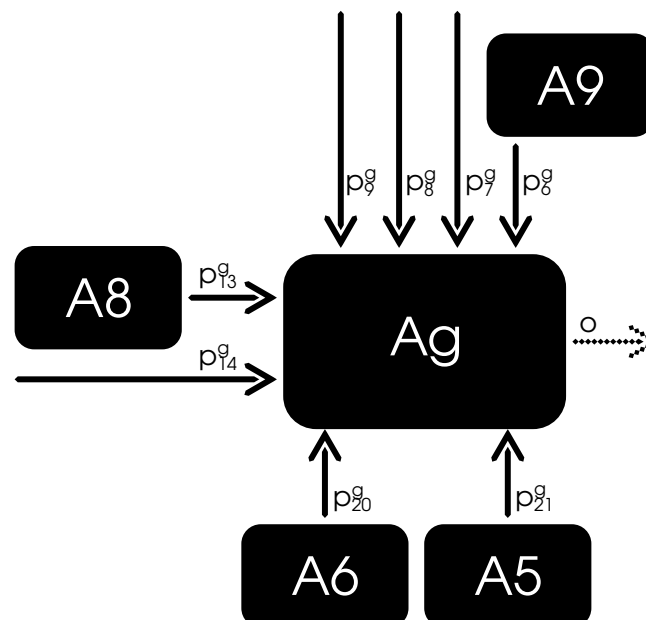


FIGURE 5 – Exemple d'activité groupée

L'enjeu du programme est donc de déterminer les valeurs p_i^g dans le cas général, quel que soit le groupement désiré et quelle que soit la base de donnée et sa structure. Dans notre exemple, $i \in \{6, 7, 8, 9, 13, 14, 20, 21\}$.

4.2 Processus d'implémentation par essai-erreur

Notons $\#$ le cardinal d'un ensemble donné.

Soit n le nombre d'activités au sein du groupe. $n = 4$ dans l'exemple présenté ci-dessus.

Soit m le nombre de montants d'échange à déterminer. $m = 8$ dans l'exemple présenté ci-dessus.

Soit I l'ensemble des arcs internes au groupe. On remarque que $\#I \geq n - 1$ et que $\#I = n - 1$ dans le cas où chaque activité du groupe n'a qu'un unique arc interne au groupe dont elle est la source. Ce n'est pas le cas dans l'exemple ci-dessus, où $A2$ possède un arc allant vers $A1$ et un arc allant vers $A3$. $I = \{p_1, p_2, p_3, p_4, p_5\}$ dans l'exemple.

Soit E l'ensemble des arcs qui vont de l'extérieur vers l'intérieur du groupe. On a par construction $\#E = m$. $E = \{p_i, i \in \{6, 7, 8, 9, 13, 14, 20, 21\}\}$ dans l'exemple.

Soit N l'ensemble des activités internes au groupe. On a par construction $\#N = n$. $N = \{A_i, i \in \{1, 2, 3, 4\}\}$ dans l'exemple.

4.2.1 À l'ordre 1

L'implémentation spontanée et naïve serait de déclarer :

$$\forall p_i \in E, p_i^g = p_i$$

Ce qui revient à copier-coller les flux initiaux à l'entrée de l'activité groupée.

Problème

Bien entendu, cette méthode ne fonctionne pas. Ou plus précisément elle ne fonctionne que dans le cas où $n = 1$. En effet, il faut prendre en compte le montant des échanges internes au groupe. La sortie du groupe (o dans l'exemple) correspond à la sortie de son activité terminale ($A1$ dans l'exemple) le montant des flux en entrée de $A1$ sont donc bien définis pour une unité en sortie de $A1$, et donc pour une unité en sortie de Ag .

En revanche, les flux en entrée de $A2$, $A3$ et $A4$ sont définis pour une unité de sortie de ces activités. Or, au sein du groupe, ce qui nous intéresse, ce n'est pas la production d'une unité de $A2$, de $A3$ ou de $A4$, mais bien p_2 unité de $A2$ et p_1 unité de $A3$.

Donc avec la méthode du copier-coller, seuls les flux p_6^g et p_{21}^g auront une valeur juste, à moins que par miracle les valeurs de p_1 , p_2 , p_3 , p_4 et p_5 ne se coordonnent parfaitement, ce qui n'est en réalité jamais le cas.

4.2.2 À l'ordre 2

Pour mieux comprendre ce qui suit, complétons un peu la notation. Un montant d'échange p_i sera noté ainsi :

$$\begin{matrix} A_{source} \\ A_{cible} \end{matrix} p_i$$

Cette notation permet de visualiser toutes les informations relatives à un arc du graphe associé au montant qui nous intéresse.

Le problème constaté précédemment peut être résolu par une simple somme de règles de 3 :

Soit $A_j \in N$ telle que A_j non terminale.

Soit $A^j p$ l'ensemble des flux sortant de A_j .

Soit ${}_{A_j}E$ la réduction de l'ensemble E aux flux ciblant A_j .
 Notons $\mathbb{1}$ la fonction indicatrice.

$$\forall {}_{A_j}p_i \in {}_{A_j}E; p_i^g = \sum_{{}_{A_j}p_k \in {}_{A_j}p} \mathbb{1}_I({}_{A_j}p_k) {}_{A_j}p_i {}_{A_j}p_k$$

Dans le cas simple et courant où un seul flux interne ${}_{A_j}p_k$ a A_j comme source, la formule donne :

$$\forall {}_{A_j}p_i \in E; p_i^g = {}_{A_j}p_i {}_{A_j}p_k$$

Le cas de l'activité terminale se traite de la même manière en ajoutant artificiellement un flux unitaire prenant l'activité terminale comme source et sans activité cible, que l'on considère appartenant à l'ensemble I . Ce flux est une implémentation concrète de o dans les figures 4 et 5.

L'application de cette méthode de calcul dans notre exemple donne :

$$\begin{aligned} p_6^g &= p_6 \cdot 1 \\ p_{21}^g &= p_{21} \cdot 1 \\ p_7^g &= p_7 \cdot p_1 \\ p_8^g &= p_8 \cdot p_2 + p_8 \cdot p_3 \\ p_9^g &= p_9 \cdot p_2 + p_9 \cdot p_3 \\ p_{20}^g &= p_{20} \cdot p_2 + p_{20} \cdot p_3 \\ p_{13}^g &= p_{13} \cdot p_4 + p_{13} \cdot p_5 \\ p_{14}^g &= p_{14} \cdot p_4 + p_{14} \cdot p_5 \end{aligned}$$

Problème

Cet algorithme fonctionne à l'ordre 2, mais ne fonctionne plus au-delà.

En effet, regardons de plus près $A4$. Il faut p_4 unités de $A4$ pour produire une unité de $A2$, or ce qui nous intéresse est une unité de $A1$, et non une unité de $A2$. Il faut donc procéder récursivement, afin de remonter le graphe d'activités jusqu'à l'activité terminale du groupement.

Si l'on ignore pour l'instant le cycle p_5 que nous traiterons plus tard, il faut non pas p_4 unités de $A4$ pour une unité de $A1$, mais bien $p_4 p_2 + p_4 p_3 p_1$ unités de $A4$ pour produire une unité de $A1$.

4.2.3 À l'ordre n

Pour résoudre le problème de l'algorithme précédent, une simple fonction récursive fonctionne, en voici un exemple non optimisé⁵ :

5. Cet algorithme manque de façon évidente d'optimisation, mais il donne une idée simple d'implémentation. De plus, le décorateur `@cache` du package `functools` de python permet déjà de résoudre une partie de ces problèmes d'optimisation. Enfin, en réalité, la taille n des groupement est toujours très faible et le temps perdu avec un algorithme peu optimisé comme celui-ci est négligeable face au temps dont a besoin `Brightway2` pour faire appel à ses backend `SQLite`.

Algorithme 1 Calcule le montant total nécessaire d'une activité interne au groupe

Entrée $A_j \in N$; $A_j p$ l'ensemble des flux sortant de A_j

Sortie La quantité p de A_j utilisée en interne du groupe

```

1 : procédure TOTALACTIVITE( $A_j, A_j p$ ) :
2 :   Si  $A_j$  est terminale alors
3 :     renvoyer 1
4 :   Sinon
5 :      $val \leftarrow 0$ 
6 :     Pour  $A_k p_k \in A_j p \cap I$ , faire :
7 :        $val \leftarrow val + \text{TOTALACTIVITE}(A_k, A_k p) \times \frac{A_j}{A_k p_k}$ 
8 :     Fin pour
9 :     renvoyer  $val$ 
10 :  Fin si
11 : Fin procédure

```

Si l'on ignore la boucle p_5 , cet algorithme donne pour chaque activité interne au groupe une quantité de pondération des flux entrants suivante :

$$\begin{aligned}
 A1 &: 1 \\
 A2 &: p_2 A1 + p_3 A3 = p_2 + p_3 p_1 \\
 A3 &: p_1 A1 = p_1 \\
 A4 &: p_4 A2 = p_4 p_2 + p_4 p_3 p_1
 \end{aligned}$$

Problème

Cette façon de faire fonctionne très bien à condition qu'il n'y ait pas de cycle interne au groupe. Il est assez facile de comprendre le problème des cycles sur un algorithme récursif comme celui-ci : tout d'abord il nécessite la mise en place d'un seuil de coupure arbitraire. Ensuite, soit le cycle est absorbant et sera asymptotiquement ignoré par l'algorithme, soit le cycle diverge et la valeur de sortie de l'algorithme divergera également.

Comme nous l'avons vu dans l'exemple, il faut ignorer le cycle p_5 pour avoir un résultat cohérent.

4.2.4 Les cycles

Pour régler cette question de cycle, il faut tout simplement revenir à la base de la logique de la structure de l'ACV : l'analyse entrée-sortie (voir section 2.1).

En effet, les valeurs calculées récursivement précédemment (la quantité de chaque activité interne au groupe) correspondent exactement aux valeurs calculées par la matrice de Leontief inverse, au niveau du secteur de sortie $A1$. Il suffit donc de calculer cette matrice, puis de multiplier sa colonne correspondant à $A1$ par les flux appartenant à E , et le tour est joué.

Dans notre exemple :

$$A = \begin{array}{c} A1 \\ A2 \\ A3 \\ A4 \end{array} \begin{array}{cccc} A1 & A2 & A3 & A4 \\ \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ p_2 & 0 & p_3 & 0 \\ p_1 & 0 & 0 & 0 \\ 0 & p_4 & 0 & p_5 \end{array} \right] \end{array}$$

$$I - A = \begin{array}{c} A1 \\ A2 \\ A3 \\ A4 \end{array} \begin{array}{cccc} A1 & A2 & A3 & A4 \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ -p_2 & 1 & -p_3 & 0 \\ -p_1 & 0 & 1 & 0 \\ 0 & -p_4 & 0 & 1 - p_5 \end{array} \right] \end{array}$$

$$L = (I - A)^{-1} = \begin{array}{c} A1 \\ A2 \\ A3 \\ A4 \end{array} \begin{array}{cccc} A1 & A2 & A3 & A4 \\ \left[\begin{array}{cccc} L_{A1,A1} & L_{A1,A2} & L_{A1,A3} & L_{A1,A4} \\ L_{A2,A1} & L_{A2,A2} & L_{A2,A3} & L_{A2,A4} \\ L_{A3,A1} & L_{A3,A2} & L_{A3,A3} & L_{A3,A4} \\ L_{A4,A1} & L_{A4,A2} & L_{A4,A3} & L_{A4,A4} \end{array} \right] \end{array}$$

Avec L_{A_j,A_k} la quantité de l'activité A_j nécessaire à la production d'une unité de A_k .
Il faut ensuite construire la matrice des flux externes⁶ :

$$F = \begin{array}{c} 6 \\ 21 \\ 8 \\ 9 \\ 20 \\ 7 \\ 13 \\ 14 \end{array} \begin{array}{cccc} A1 & A2 & A3 & A4 \\ \left[\begin{array}{cccc} p_6 & 0 & 0 & 0 \\ p_{21} & 0 & 0 & 0 \\ 0 & p_8 & 0 & 0 \\ 0 & p_9 & 0 & 0 \\ 0 & p_{20} & 0 & 0 \\ 0 & 0 & p_7 & 0 \\ 0 & 0 & 0 & p_{13} \\ 0 & 0 & 0 & p_{14} \end{array} \right] \end{array}$$

Enfin, il faut multiplier cette matrice par le vecteur colonne correspondant à la colonne de la matrice L associée à l'activité de sortie du groupe. Dans l'exemple, il s'agit de $A1$:

$$A1_L = \begin{bmatrix} L_{A1,A1} \\ L_{A2,A1} \\ L_{A3,A1} \\ L_{A4,A1} \end{bmatrix}$$

L'on obtient donc le vecteur p^g des flux d'entrée du groupement :

6. Cette matrice est équivalente à la matrice de biosphère B de la section 2.2.3, dans le cadre d'une ACV complète. En effet, un groupement d'activité est équivalent à une ACV dont les flux de biosphère seraient tous les flux allant de l'extérieur vers l'intérieur du groupe.

$$p^g = F \cdot A1_L = \begin{bmatrix} p_6^g \\ p_{21}^g \\ p_8^g \\ p_9^g \\ p_{20}^g \\ p_7^g \\ p_{13}^g \\ p_{14}^g \end{bmatrix}$$

Problème

L'algorithme présenté ici est fiable, comme montré dans la section 4.3. Cependant, il reste un dernier cas à traiter : les co-produits. Comme expliqué dans un encadré de la section 2.2.2, les co-produits des activités de la base de donnée sont définis formellement par des entrées négatives de l'activité. Insérées telles quelles dans la matrice A , les jeux de test nous indiquent que ces données négatives génèrent des erreurs.

4.2.5 Les entrées négatives

Pour comprendre que faire des entrées négatives, il faut revenir à leur signification : "Pour produire une unité d'une activité, celle-ci besoin d'une certaine quantité de service de traitement du déchet concerné". Ce service de traitement des déchets aura ensuite ses propres entrées, et donc ses propres impacts. Du point de vue des relations entrées-sorties représentées par A , il s'agit donc bien d'une entrée positive.

Il faut donc, au moment de la construction de A , prendre la valeur absolue des flux internes, ce qui donne dans notre exemple :

$$A = \begin{matrix} & A1 & A2 & A3 & A4 \\ \begin{matrix} A1 \\ A2 \\ A3 \\ A4 \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ |p_2| & 0 & |p_3| & 0 \\ |p_1| & 0 & 0 & 0 \\ 0 & |p_4| & 0 & |p_5| \end{bmatrix} \end{matrix}$$

4.3 Preuve de fonctionnement

4.3.1 Invariant

Pour pouvoir déterminer si le programme est fonctionnel, il faut identifier un invariant qu'il est possible de tester. En l'occurrence, l'invariant qui s'impose est l'ICV. En effet, l'ICV d'une unité de l'activité groupée est théoriquement strictement égale à l'ICV d'une unité de l'activité de sortie du groupe avant groupement. Le vecteur g de l'activité groupée doit donc être terme à terme égal au vecteur g de l'activité de sortie du groupement (voir section 2.2.3).

4.3.2 Approche de la démonstration

L'égalité des ICV se démontre en travaillant avec des matrices de technosphere et de Leontief inverse par blocs. En effet, la matrice interne A d'un groupe correspond à un sous-bloc de la matrice de technosphere de l'ensemble de la base de données (aux permutations près).

Le mérite de la démonstration revient à Jean-Yves COURTONNE. Pour ne pas prendre trop de place dans le corps du rapport, celle-ci est en annexe A.

5 Tests

Afin de s'assurer que l'implémentation du programme couvre bien tous les cas souhaités, et qu'elle ne comporte pas d'erreurs, il faut pouvoir la tester.

5.1 Système de test

Comme vu dans la section 4.3, pour que le groupement se soit bien effectué, il faut que sa remontée d'ICV soit égale à celle de son activité terminale avant le groupement.

5.1.1 Test unitaire

Un test unitaire est ici un fichier contenant une série d'activités à grouper⁷.

Le système de test va alors importer le fichier, grouper les activités, et effectuer les remontées d'ICV pour l'activité terminale d'une part et pour le groupe d'autre part. Il obtient ainsi deux vecteurs de flux de biosphère : g^{act} et g^{grp} . Bien entendu $\#g^{act} = \#g^{grp}$, l'on notera cette valeur $\#g$.

Il vérifie ensuite que $\forall i \in \{1, \dots, \#g\}; g_i^{act} = g_i^{grp}$.

En théorie, si c'est le cas, le test passe, sinon non.

5.1.2 Jeux de test

Les jeux de test sont issus des activités d'Écoinvent. Certains sont construits à la main, mais une bonne partie de ceux-ci est également générée aléatoirement, ce qui permet de découvrir des cas particuliers qui ne sont pas forcément évidents à identifier en amont.

Le système lance ensuite tous les fichiers de test présents au sein d'un dossier spécifique, et s'assure qu'ils passent tous la vérification unitaire.

Si tous les tests passent la vérification unitaire, le programme et son implémentation sont considéré-e-s fiables.

Un point fort avec cette façon de fonctionner est la possibilité d'intégrer le système de test directement à l'usage du programme. En effet, il est possible de s'assurer, au moment où l'on utilise le programme pour grouper des activités, du test direct du groupement généré pour vérifier qu'il est, de même que le jeu de données qui en découle, fiable.

5.2 Problématique technique

Une fois le système de test décrit ci-dessus implémenté, son utilisation nous place face à un problème. Même si mathématiquement cette façon de tester est idéale, dans la pratique, des divergences apparaissent systématiquement dans les valeurs, notamment dues à la précision limitée des ordinateurs. Pire, si un test unitaire donné est effectué avant ou après un autre test unitaire indépendant, il ne donnera pas le même résultat. En d'autres mots, rajouter un test dans le jeu peut avoir, de façon aléatoire, pour conséquence qu'un autre test ne passe soudainement plus la vérification.

⁷ La première activité indiquée par le fichier doit être l'activité terminale du groupe et l'ensemble des activités du fichier doivent être assurément conformes aux propriétés imposées pour les sous-graphes de groupement (par exemple la connexité).

5.2.1 Explication

Cette divergence au sein de tests théoriquement justes s'explique par l'analyse numérique. En effet, la remontée d'ICV se fait via l'arithmétique suivante (voir section 2.2.3) :

$$g = BA^{-1}f$$

Or, la matrice A est d'une taille égale au nombre d'activités dans la base de données (par construction), soit presque 15000 dans le cas d'Ecoinvent 3.4. Elle est de plus extrêmement creuse et possède des valeurs ayant des ordres de grandeur différents, pouvant parfois aller de 10^{-13} à 10^{11} .

Cette remarque vaut également pour la matrice B , mais dans ce cas-ci, le problème est moindre pour deux raisons :

1. Seule une des deux dimensions de la matrice B est aussi importante que celle de la matrice A , l'autre correspond au nombre d'interactions avec l'environnement, soit presque 10 fois moins dans le cas qui nous intéresse.
2. L'arithmétique est moins contraignante envers cette dernière, puisqu'il s'agit de simplement la multiplier à droite, tandis que la matrice A nécessite une inversion, laquelle implique des divisions.

Le cœur de la problématique se trouve donc au niveau de la matrice A . Grâce à sa construction, celle-ci est extrêmement sensible non seulement à la propagation d'erreur - notamment celle des erreurs numériques -, mais également susceptible d'en générer de non-négligeables lors de son inversion.

La divergence entre g^{act} et g^{grp} est donc liée à des variations théoriquement indépendantes de la matrice A . En effet, g^{act} est calculée avec une matrice A d'une taille donnée. Puis l'activité groupée est enregistrée dans la base de données, augmentant la taille de A de 1, faisant diverger le résultat de l'opération d'inversion de A (à cause de son instabilité), et donnant un vecteur g^{grp} différent de g^{act} .

Précision sur une approche triviale de résolution

Cette façon de présenter le problème donne l'impression qu'il suffit d'ajouter le groupe à la base de donnée avant d'effectuer le calcul, et alors A^{-1} sera identique dans les deux cas.

En réalité, il s'agit bien de ce qui est fait dans le système de test, mais l'activité groupée est enregistrée dans une base de données à part (voir figure 3), pour éviter d'interférer avec les données officielles d'Ecoinvent, tout simplement pour des raisons pratiques de conservation de la distinction entre les données de référence et les données calculées.

Or **Brightway2** va détecter si les données sont partagées entre plusieurs bases et va concaténer ces dernières le cas échéant. Dans le cas de référence, A correspond à Écoinvent uniquement, et dans le cas calculé, A correspond à la concaténation d'Écoinvent et de la base de données des activités groupées.

De plus, même dans le cas théorique où A^{-1} n'est pas recalculée, des imprécisions persisteront au niveau du reste de l'arithmétique, et du calcul des flux entrant des groupes d'activités (voir section 4). Il n'y a donc jamais égalité exacte des résultats, même si les divergences sont réduites.

Conditionnement matriciel

Le conditionnement d'une matrice [36] est une valeur supposée représenter son instabilité, c'est-à-dire sa sensibilité aux erreurs en entrée, particulièrement dans le cas où la matrice doit

être inversée. Une matrice mal conditionnée va voir son inverse diverger fortement si de légères perturbations viennent modifier les données en entrée⁸.

Il existe plusieurs indicateurs du conditionnement d'une matrice inversible. Un indicateur imparfait, mais frappant, est la valeur de son déterminant. S'il est très grand ou très petit, l'intuition est que la matrice est mal conditionnée⁹.

Calculons donc $\det(A)$ dans notre cas avec la base de donnée d'Écoinvent. La fonction `python numpy.linalg.det` renvoie `-inf...` D'après sa documentation, ce résultat est probablement dû à des dépassements de la représentation flottante des nombres réels lors du calcul. Il est alors possible d'utiliser la fonction `numpy.linalg.slogdet` qui calcule le logarithme du déterminant, et qui elle renvoie environ `(-1, -46)`, ce qui signifie que le déterminant est de l'ordre de grandeur de -10^{-20} .

Au delà de la valeur du déterminant, le conditionnement d'une matrice inversible est défini par une valeur notée κ , qui pour une norme $\|\cdot\|$ donnée vaut :

$$\kappa(A) = \|A^{-1}\| \|A\|$$

Le conditionnement d'une matrice est considéré comme bon si κ est proche de 1.

Cette valeur peut être calculée avec `numpy` via la fonction `numpy.linalg.con`, et qui donne dans notre cas un ordre de grandeur de 10^{23} , sachant que l'ajout d'une unique activité dans la base de données peut faire varier la valeur de κ d'environ son double.

En conclusion, la matrice A est effectivement très mal conditionnée et le problème vient a priori de là.

5.2.2 Approches de résolution

Il est possible d'appréhender le problème de 3 façons :

1. Considérer qu'il s'agit d'un problème intrinsèque à la méthodologie de l'ACV, et entrer dans un fatalisme frôlant l'abandon total de foi en la science.
2. Essayer de limiter les erreurs numériques.
3. Déterminer des indicateurs et des seuils associés pour discriminer les erreurs relevant de l'erreur numérique et les erreurs relevant de bugs dans l'implémentation de l'algorithme.

Bien entendu, lors de ce stage, nous ne nous arrêterons pas sur la première approche, même si celle-ci mériterait d'être creusée par la discipline : je n'ai pas identifié de littérature sur cette question-ci précisément.

Il est en revanche pertinent d'explorer les deux autres approches.

Limitation des erreurs numériques

Il y a plusieurs approches pour limiter les erreurs numériques. L'une d'entre elles consiste à essayer de préconditionner la matrice, c'est-à-dire à améliorer le κ de cette dernière. Une autre serait d'effectuer les tests du programme sur une base de donnée artificielle et bien conditionnée.

8. La page Wikipédia sur le conditionnement propose un exemple de matrice mal conditionnée.

9. Cette façon d'estimer le conditionnement d'une matrice est insuffisante, par exemple dans le cas d'une matrice diagonale avec uniquement des valeurs très faibles, mais très proche, le déterminant va être mauvais, mais le problème est résolu par un simple changement d'échelle de la diagonale [37].

Bien entendu, ce second cas est compliqué à mettre en place, et ne donne pas la certitude que tous les cas réels peuvent être atteints. Il est alors intéressant de creuser la question du préconditionnement des matrices qui nous intéressent.

Le préconditionnement matriciel est un champ de recherche à part entière et ce problème demanderait un stage entier à lui tout seul pour explorer les différentes méthodes. J'ai donc exploré une méthode bien spécifique, mais il existe probablement de nombreuses façons de faire, davantage adaptées, que je n'ai pas pu étudier faute de temps.

Cette méthode pour rendre l'algèbre moins instable consiste à changer l'unité des lignes des matrices A et B . En effet, les représentations des nombres flottants sont plus dense autour des nombres ayant 1 comme valeur absolue : l'utilisation de nombres trop grands ou trop petits mène à des approximations lors des calculs, pouvant mener à des erreurs importantes. Un cas d'école est la situation suivante :

$$\frac{1}{x + \epsilon - x}$$

Avec $|x| \gg |\epsilon|$, $x + \epsilon \simeq x$ et ϵ est alors négligé. Le calcul donne donc un dénominateur nul, et donc une division par 0.

L'objectif est donc d'équilibrer les valeurs des lignes des matrices autour de 1. La façon la plus simple d'implémenter ce type de méthode est de centrer le maximum absolu et le minimum absolu (non nul) de la ligne par une simple multiplication de cette dernière par un facteur λ .

Soit, pour une ligne l donnée de la matrice à équilibrer, M_l son maximum en valeur absolue et m_l son minimum non nul en valeur absolue. L'objectif est qu'après multiplication par un même facteur λ_l , ces deux valeurs soient équilibrées en terme d'ordre de grandeur, ce qui se traduit par le système :

$$\begin{cases} \lambda_l M_l = 10^a \\ \lambda_l m_l = 10^{-a} \end{cases} \\ \Rightarrow \lambda_l^2 M_l m_l = 1 \\ \Leftrightarrow \lambda_l = \frac{1}{\sqrt{M_l m_l}}$$

Il suffit alors de construire la matrice diagonale P_{mat} qui permettra d'équilibrer la matrice (mat) qui nous intéresse, et dont l'inverse permettra de revenir aux valeurs réelles de résultat :

$$P_{mat} = \begin{pmatrix} \frac{1}{\sqrt{M_1 m_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{M_2 m_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{M_n m_n}} \end{pmatrix} \quad P_{mat}^{-1} = \begin{pmatrix} \sqrt{M_1 m_1} & 0 & \dots & 0 \\ 0 & \sqrt{M_2 m_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{M_n m_n} \end{pmatrix}$$

Au lieu de calculer A^{-1} , il faut donc calculer $P_A^{-1}(P_A A)^{-1}$. De même pour la matrice B . On obtient donc :

$$g = P_B^{-1}((P_B B)(P_A^{-1}(P_A A)^{-1}))f$$

Malheureusement, cette méthode s'est montrée peu effective pour trois raisons :

1. Cette manière de construire les facteurs est un peu trop grossière telle quelle, et par conséquent empirait parfois le conditionnement de la matrice.
2. Le fait de recentrer de toutes petites valeurs sur les lignes, faisait parfois au contraire apparaître des erreurs non négligeables là où elles étaient négligeables auparavant, notamment avec le préconditionnement de B .
3. La structure technique de `Brightway2` ne permettait pas de facilement accéder à la valeur de A^{-1} , en raison de l'utilisation de solveurs faisant appel à des structures de données spécialisées pour des raisons d'optimisation.

La tentative de préconditionnement des matrices a donc été abandonnée, mais en gardant à l'esprit qu'il s'agit d'une piste parfaitement viable, le stage n'étant cependant pas assez long pour creuser cette voie.

Indicateurs et seuils

Le conditionnement matriciel étant une direction trop longue et trop complexe à explorer dans le cadre de ce stage, il fallait quand même réussir à déterminer si les valeurs calculées par le programme étaient effectivement fiables. Pour cela, il fallait déterminer des indicateurs statistiques sur les résultats du test et des seuils associés à partir desquels le résultat du test était considéré comme non acceptable.

À partir des deux vecteurs g^{act} et g^{grp} , il est possible de récupérer un certain nombre d'indicateurs de comparaison. Personnellement, j'en ai retenu deux grandes catégories : les statistiques sur les erreurs relatives et les statistiques sur les erreurs absolues.

Il faut donc construire deux nouveaux vecteurs : g^{abs} et g^{rel} de la façon suivante :

$$\begin{aligned} \forall i \in \{1, \dots, \#g\}, \\ g_i^{abs} &= |g_i^{act} - g_i^{grp}| \\ g_i^{rel} &= \frac{|g_i^{act} - g_i^{grp}|}{|g_i^{act}|} \end{aligned}$$

Ces vecteurs sont les vecteurs d'erreurs absolues et d'erreurs relatives. Il est ensuite facile d'effectuer des statistiques sur chacun de ces vecteurs : moyenne, maximum, etc.

Certaines statistiques s'effectuent en dehors de ces vecteurs, mais elles sont minoritaires. C'est le cas du nombre de singularités (expliqué ci-dessous).

À l'heure actuelle, ont été retenus les indicateurs suivants :

- Le maximum des erreurs absolues et celui des erreurs relatives.
- Le minimum des erreurs absolues.
- La p-value du test du χ^2 de la loi de Benford pour les deux vecteurs.
- Le nombre de fois où un élément de l'un des deux vecteurs est nul tandis que la valeur de l'autre vecteur est non négligeable. J'ai nommé cet indicateur le nombre de singularités¹⁰.

¹⁰. Notons que les singularités empêchent toute utilisation de leur erreur relative, celle-ci n'étant pas définie en 0.

Apparté sur la loi de Benford

L'usage de la loi de Benford dans ce genre de cas est une simple initiative personnelle, je n'ai pas trouvé de littérature à ce sujet.

Pourtant, qualitativement, j'ai le sentiment que cette approche a du potentiel dans certains cas spécifiques. On remarque d'ailleurs que quand elle est utilisée sur le vecteur d'erreur absolue, elle est très fiable : trop même, puisqu'elle ne dévie pas dans la plupart des cas, même si un bug important est présent dans le code. Cette analyse sur le vecteur d'erreurs relatives en revanche est soumise au problème inverse : le suivi de la loi de Benford par le vecteur d'erreurs relatives est trop instable pour identifier de manière claire si un problème apparaît. Cela se visualise bien avec les jeux de graphes de l'annexe C.

Dans tous les cas, cet indicateur permet d'identifier une anomalie, en aucun cas d'assurer de manière certaine que l'algorithme n'est pas fiable. Pour en comprendre mieux les mécanismes, cette approche de l'intégration continue mériterait d'être étudiée en profondeur par l'analyse numérique, pour identifier si elle est pertinente ou non. Cette question n'est pas mon sujet de stage, donc tant pis. Mais elle est intéressante.

Finalement, l'indicateur qui, après implémentation, semble avoir le plus fait ses preuves est le maximum d'erreur relative.

Il faut également noter qu'un seuil de coupure est défini (accessible via le fichier de configuration du programme), en dessous duquel il est considéré que la valeur est négligeable. Si les deux valeurs à comparer sont sous ce seuil, elles sont ignorées, sinon non, à moins qu'il ne s'agisse d'une singularité, laquelle est comptabilisée à part et n'apparaît donc pas dans les vecteurs g^{abs} et g^{rel} .

L'ensemble des indicateurs calculés dans le programme apparaissent dans le tableau de l'annexe B. La plupart n'ont pas été paramétré pour la détection d'erreurs, mais permettent une vérification qualitative - car sauvegardés dans les logs - de la fiabilité des données dans le cas où la vérification automatique aurait échoué. Parmi les logs en question, se trouvent également des sauvegardes des vecteurs g^{abs} et g^{rel} , un ensemble d'informations relatives au graphe du groupe, les graphiques de répartition selon la loi de Benford et un graphique des valeurs calculées en fonction des valeurs de référence.

Fixation du déterminisme des résultats

Finalement, un dernier ajustement à la marge a besoin d'être mis en place, pour éviter le cas où l'ajout d'un test dans le jeu ne déstabilise le résultat des autres tests. Afin d'éviter ce problème, il suffit de supprimer l'activité générée par chaque test une fois celui-ci effectué. Ainsi, la taille de la base de données n'augmente pas de un à chaque test, mais reste systématiquement à la même taille, égale à la taille d'Écoinvent plus un.

Cette méthode a pour conséquence, dans une certaine mesure, de cacher la poussière sous le tapis. En effet, elle consiste à dire "Ce test passe dans une configuration spécifique de la BD, donc il faut fixer cette configuration de BD", tandis qu'il se peut que le test passe dans une configuration spécifique grâce à un coup de chance, et qu'en moyenne, il ne passe pas les autres configurations. Les problèmes soulevés par le test passeront donc sous le radar.

Ceci dit, ce cas de figure est a priori suffisamment peu probable pour ne pas vraiment avoir à s'en soucier : il semblerait que ce soit dans l'autre sens que les tests fonctionnent (d'après mon expérience personnelle), c'est-à-dire qu'une fois de temps en temps ils ne passent pas quand ils sont juste, et non pas une fois de temps en temps ils passent quand ils sont faux.

5.3 Choix finaux d'implémentation

En résumé, le système de test fonctionne de la façon suivante :

- Aucun pré-conditionnement n'est effectué sur les matrices, mais dans le cas où la matrice $I - A$ interne au groupe (voir section 4.2.4) est mal conditionnée, le logger du programme lève un warning à l'utilisateur-riche, permettant de l'avertir que le résultat du test peut être instable.
- Un ensemble de statistiques sont relevées sur les vecteurs de test (voir annexe B), chacune est vérifiée, et si certaines d'entre elles ne respectent pas les contraintes de seuils imposées, des warnings et des erreurs sont levées en fonction de la sévérité de la divergence.
- Un ensemble de logs est enregistré pour pouvoir effectuer une vérification qualitative et du debugging en cas de levée d'erreurs pour un groupement donné.
- Une vérification sur un système inspiré de la simulation de Monte Carlo [38] est mis en place pour s'assurer qu'un test ne passant pas la vérification à un moment donné, soit simplement sur une configuration de la BD le mettant à mal ponctuellement, et non pas sur une véritable erreur récurrente (voir section 5.3.1).

Ce système de test peut être effectué soit sur des jeux de données à la manière d'une intégration continue grâce à un script spécifique (`test.py`), soit exécuté à la suite de l'utilisation du programme de groupement pour s'assurer que les données groupées sont fiables. En cas de problème, une première procédure de vérification de Monte Carlo est mise en place, puis une procédure de vérification qualitative peut ensuite être mise en place grâce aux logs.

5.3.1 Vérification façon Monte Carlo

Un problème auquel il faut absolument faire face avec ce système de test, c'est le cas des situations déviantes. J'entends par là que pour un test donné, s'il est "juste", il passera la vérification dans la majorité des cas, mais échouera à la passer une fois de temps en temps. En effet, des instabilités fortes peuvent apparaître en fonction de la configuration de la base de données.

Bien entendu, comme dit précédemment, une fois que le test passe, dans le cas où la base est de taille de celle d'Écoinvent plus 1, le tour est joué puisque cette configuration de la base de données est fixée. Mais cela ne vaut que pour les jeux de test construits à la main, dans le cas de la vérification par l'utilisateur-riche de son groupement, le comportement est imprévisible.

Il s'agit donc, lorsqu'une ou plusieurs statistiques d'un test ne passent pas la vérification, de lancer ledit test plusieurs fois, en modifiant à chaque itération la configuration de la BD. Concrètement, ce changement de configuration se traduit par un gonflage artificiel de la BD.

Il s'agit ensuite d'effectuer la moyenne sur toutes les itérations de l'indicateur ayant levé une erreur initialement, et de regarder si cette moyenne passe finalement la vérification ou non. Si la moyenne passe la vérification, il semble raisonnable d'estimer qu'il s'agissait initialement d'une fausse alerte, sinon il y a matière à regarder plus en détail pourquoi le test ne passe pas.

Une question qui se pose dans ce type de situation est le choix de la moyenne à utiliser : Arithmétique ? Géométrique ? Harmonique ? Ce choix dépend de l'indicateur statistique. Dans certains cas, la moyenne arithmétique est intéressante, mais dans les cas où il est question d'ordres de grandeur très variables, la moyenne géométrique semble plus pertinente¹¹. Il s'agit de ce choix de moyenne qui est indiqué en dernière colonne du tableau de l'annexe B.

11. La moyenne géométrique peut être vue comme la moyenne des ordres de grandeur. En effet, le logarithme de la moyenne géométrique est égale à la moyenne arithmétique des logarithmes.

Cette vérification, qui ressemble à une méthode de Monte Carlo, est très gourmande en calculs, et ne peut donc pas être effectuée systématiquement. Pour cette raison, elle n'est utilisée que dans un second temps en vérification, uniquement si le test concerné a levé une erreur.

6 Bilan et suite du projet

6.1 Bilan technique du projet

Finalement, aura été implémenté durant le stage :

1. La possibilité de grouper manuellement des activités d'Écoinvent, au moins de sa version 3.4.
2. Une structure MVC, ce qui rend le développement futur facilement flexible.
3. Une interface en ligne de commande.
4. Un fichier de configuration et un ensemble d'options en ligne de commande.
5. Un système de test, y compris dans le cadre d'une vérification par l'utilisateur·rice. Il est paramétrable.
6. Un système d'enregistrement de logs.
7. Une documentation technique (en anglais) sous forme de docstrings, rédigées selon le format Google, et donc facilement exportable via un système comme `pdoc`.
8. Un README (en anglais) regroupement toutes les informations nécessaires à l'installation, à l'utilisation et au développement du programme.

6.2 Limites

La principale limite du programme est sa lenteur. En effet, `Brightway2` n'est pas rapide pour faire appel à ses backup des bases de données. L'export d'une activité groupée, et plus précisément l'export de chacun de ses échanges, est réellement lent. Cela ne pose en général que peu de problèmes, mais est embêtant lorsqu'il s'agit d'effectuer de nombreux exports à la chaîne, comme c'est le cas pour les vérifications de test par un système de Monte Carlo (voir section 5.3.1). Il existe probablement des méthodes plus avancées qui permettraient de limiter ce problème, au niveau du choix du format du backup ou des méthodes utilisées pour y faire appel : par exemple en enregistrant tous les échanges d'un coup, plutôt que chacun séparément. Je n'ai pas eu encore le temps de les explorer en profondeur, en tout cas des recherches rapides de surface n'ont donné aucun résultat. Il devrait être envisageable également d'améliorer le système de Monte Carlo afin que celui-ci n'exporte qu'une seule fois l'activité concernée au lieu de l'exporter à chaque itération, mais il faudrait alors s'assurer que le résultat reste fiable.

Une autre limite est le non préconditionnement des matrices calculées par l'algorithme (voir section 4.2.4). Il serait intéressant de faire en sorte de limiter les erreurs de calcul lorsque les échanges internes aux groupes sont trop instables, notamment avec des ordres de grandeur fortement différents. Actuellement, un simple warning est levé.

De plus, certaines fonctions et méthodes du programme y gagneraient à être optimisées, même si le gain de temps généré serait de toute façon négligeable devant le temps d'export dans les backup `SQLite`.

Encore, un problème de taille est que `Brightway2`, lors de l'import d'Écoinvent, ne garde pas en mémoire la métadonnée de durée de validité des données, or cette information est cruciale pour l'intégration dans le projet MATER des données récoltées via le programme. Jeanne BARRIOL aura dû faire ce travail à la main, ce qui ne semble pas être optimal du tout.

Enfin, Une autre donnée potentiellement importante pour la suite de MATER est perdue : les informations sur les incertitudes des données, qu'il faudrait idéalement intégrer dans les activités

groupées en fonction des incertitudes de ses activités internes. Je n'ai pas les compétences nécessaires telles quelles pour avoir eu le temps de les intégrer.

6.3 Où aller pour la suite ?

Durant mon stage, je n'aurai pas eu le temps d'implémenter l'ensemble des éléments qui intéressaient l'équipe d'ISTerre¹², voici quelques éléments et pistes d'approche pour la suite :

6.3.1 Déploiement

L'objectif in fine de ce programme était d'être déployé sur un serveur possédant une interface web et une API en ligne. Ce déploiement était prévu via **Docker** (voir section 3.2.2).

Faute de temps¹³, cette partie du projet n'a pas encore été mise en place.

La structure du logiciel a cependant été réfléchi en conséquence, puisque le modèle MVC est particulièrement adapté aux applications web. Certains ajustements auront certainement besoin d'être fait au niveau du contrôleur, mais globalement seule l'interface reste à faire : j'envisageais d'utiliser **Flask**, cette librairie étant la plus populaire pour le développement web en **Python**. Le module `src.interface` y est dédié, notamment le fichier `app.py` actuellement vide.

6.3.2 Interface directe avec le code de MATER

J'avais envisagé l'interface avec le code **python** de **MATER** via une API, développée en parallèle de l'application web. Le module `src.API` était supposé y être dédié.

Sinon, une interface directe est envisageable, via l'intégration de tableaux **pandas**, la structure de données utilisée pour l'implémentation du projet **MATER**.

6.3.3 Groupement semi-automatique

A priori, il ne parait pas évident d'effectuer des groupements automatiques. En tout cas, nous n'avons pas trouvé beaucoup d'invariants qui pourraient permettre d'automatiser les groupements.

En revanche, il est envisageable de grouper de façon semi-automatique une partie des activités. En effet, par exemple, on trouve régulièrement dans la BD des "market activities", dont certaines ont comme entrée plusieurs fois la même activité, mais provenant de différentes localisations. Il faut alors remonter la chaîne dans chaque cas, tandis que l'unique variante d'un cas à l'autre est la localisation.

Il serait assez facile, modulant une amélioration du contrôleur, que l'utilisateur-riche puisse indiquer plusieurs activités similaires à un moment donné, puis que le système détecte ensuite les branchements similaires à ceux déclarés sur une branche donnée, lorsqu'ils existent.

6.3.4 Identification des priorités d'intégration dans MATER

Il est déjà possible d'identifier à la main les facteurs d'impact d'une activité groupée les plus importants, tout simplement en effectuant l'ACV de ladite activité. Cette opération est facile à

12. Au moment de la rédaction de ce rapport de stage, il me reste un mois et demi de stage à venir, peut-être que certains éléments présentés ici auront évolué avant mon départ.

13. Le système de test a pris bien plus de temps que prévu à développer.

effectuer, puisque toutes ces données sont accessibles via l'interface graphique de **Brightway2** : Activity-Browser.

Il serait cependant intéressant de l'intégrer automatiquement dans le système de groupement, en déclarant notamment quels branchements sont déjà présents dans MATER et quels branchements ne le sont pas, puis, que le programme indique quelles sont les données avec le plus fort impact qui ne sont pas déclarées dans MATER.

6.3.5 Données prédictives

La modélisation MATER nécessite énormément de données, et notamment des données sur des impacts environnementaux des technologies du futur. Ces données peuvent être simulées dans les bases de données d'ACV, grâce à une technologie appelée **Wurst** [27]. Une (plusieurs ?) équipe de recherche s'est par exemple déjà servie de cette technologie pour effectuer des ACV prédictives de scénarios de transition [25].

Il faudrait alors modifier les données de la BD avant groupement selon des critères prédictifs d'évolution d'un certain nombre de facteurs, puis grouper les activités en fonction et en tirer des données pour les années non renseignées.

Une approche plus sûre, bien que d'avantage limitée, serait d'effectuer également les groupements avec les données des différentes versions d'Écoinvent, qui sont mises à jour au cours du temps. Cette approche permettrait d'avoir des données plus fiables sur les années passées, mais il faudra de toute façon estimer les évolutions futures des technologies.

7 Évaluation de l'impact socioenvironnemental du projet

Il est demandé, dans le cadre de la rédaction du rapport de stage, d'effectuer une analyse de l'impact social et écologique de ces 6 mois de projet de fin d'études.

Tout d'abord, commençons par rappeler que l'écologie n'est pas l'environnementalisme [39] : la première s'intéresse à la pensée systémique des causes profondes des problèmes environnementaux, le second s'intéresse à des ajustements à la marge ayant pour but de limiter l'impact du système politico-économique actuel sur son environnement. Essayons ici de faire de l'écologie.

L'écologie est intimement liée aux questions sociales [40] et politiques [41]. Elle ne peut, par essence, pas se réfléchir en dehors du cadre global de la société. Nombre de ces combats sociétaux sont liés. Certain-es auteur-ices défendent d'ailleurs l'idée que tout n'est qu'une question de rapports de domination [42]. Même le concept de "nature" n'est pas considéré une constante anthropologique [43].

S'il y a cependant bien une certitude, il ne fait aucun doute qu'il s'agisse du constat que la question est complexe, transdisciplinaire et pas prête d'être solutionnée malgré l'urgence.

7.1 Bilan personnel

Commençons par une petite prise de recul personnelle sur mes impacts environnementaux en tant qu'individu dans le cadre professionnel, bien que cette approche-ci ne me semble pas être la plus pertinente. En effet, on retrouve des cas de plus en plus nombreux d'ingénieur-e-s déserteur-ses : citons par exemple le discours de remise de diplôme d'AgroParisTech de 2022 [44].

Ceci dit, j'ai travaillé durant mon stage sur l'analyse des impacts environnementaux, c'est donc l'occasion de montrer que des indicateurs chiffrés existent. Tant qu'à y être, j'ai donc fait une ACV grossière de mon stage, en voici le cadre :

- La pollution induite par l'utilisation de mon ordinateur portable, en considérant qu'il fait $3.5kg$ et que sa batterie pèse $1kg$. Cet ordinateur a 6 ans et devrait bien survivre encore au moins 2 ans. Le stage est de 6 mois (24 semaines), à raison d'une utilisation de l'ordinateur de $7h$ par jour, 5 jours par semaine. Ce PC tire une puissance d'environ $10W$ en moyenne d'après le fichier système `/sys/class/power_supply/BAT0/power_now`. N'est pas pris en compte la pollution induite par l'usage d'internet durant le stage, qui est, a priori, loin d'être négligeable.
- La pollution induite par les trajets domicile-travail : sur 24 semaines, à raison de 4 jours par semaine (le 5^{ème} étant en télétravail), la moitié étant effectuée en bus, l'autre moitié étant effectuée à vélo. J'habite à environ $10km$ de mon lieu de travail.
- La pollution induite par mon seul déplacement professionnel : un aller-retour en train de Grenoble à Clermont-Ferrand ($300km$ environ).
- Ma consommation de café, à hauteur de deux expressos ($40ml$) par jour, 4 jours par semaine.

Ne sont pas pris en compte :

- Les repas du midi (mais bon, je suis végétarien, on va dire tant pis, et puis même si je ne travaillais pas, je mangerais¹⁴).

14. Il y a d'ailleurs probablement une synergie positive ici : je mange moins bien chez moi qu'au restaurant d'entreprise et l'impact individuel est réduit au niveau des coûts constant lors de l'usage de structures collectives.

- L'utilisation des bâtiments¹⁵.
- Plein d'autres choses, comme les effets rebonds, notamment ceux du télétravail.

Les données sont vieilles, puisque tirées d'Écoinvent 3.4, presque aucune n'est régionalisée, ni dimensionnée spécifiquement : j'ai notamment pris l'activité de fabrication d'ordinateur portable par défaut de la BD, sans prendre en compte les spécificités réelles de mon PC. Il est donc notamment inutile d'essayer de faire une analyse de Monte Carlo sur les résultats : les données sont bien trop lacunaires.

Les facteurs d'impacts sélectionnés sont les suivants [45] :

Catégorie d'impact	Méthode associée
Climate change, GWP100a	IPCC 2007
Material Ressources	CML v4.8 2016
Land occupation	ReCiPe
Ecotoxicity	USEtox
Human toxicity	USEtox

TABLE 4 – Liste des impacts et de leur méthode dans le cadre de l'ACV du stage

Au niveau des catégories d'impacts, les top 3 (et plus si pertinent) des résultats sont les suivants :

Changement climatique

1. Transport quotidien en bus : 56%
2. Trajet en train : 26%
3. Trajets quotidiens à vélo : 7%

Sur un total de 184kgeqCO₂.

Usage des sols (urbain et agricole)

1. Graines de café : 61%
2. Transport quotidien en bus : 22%
3. Trajet en train : 11%

Sur un total de 0.618especies/m².

Toxicité des écosystèmes

1. Fabrication de l'ordinateur portable 36%
2. Trajets quotidiens à vélo : 17%
3. Transport quotidien en bus : 17%
4. Graines de café : 16%

Sur un total de 1220CTUe.

Santé humaine

15. L'INRIA Montbonnot est chauffé au gaz, son bilan carbone est donc très mauvais à ce niveau.

1. Fabrication de l'ordinateur portable 45%
2. Trajets quotidiens à vélo : 15%
3. Trajet en train : 14%

Sur un total de $1.68 \cdot 10^{-5} CTUh$.

Consommation en ressources

1. Électricité de fonctionnement de l'ordinateur : 38%
2. Trajet en train : 31%
3. Fabrication de l'ordinateur : 16%

Sur un total de $6.88 \cdot 10^{-7} kgeqSb$.

Plusieurs éléments peuvent être constatés :

- Un secteur d'impact prédominant : les transports, et plus spécifiquement l'usage du bus.
- Un autre secteur d'impact prédominant du côté de la toxicité : la fabrication de l'ordinateur avec lequel je travaille.
- Tout ce qui a attiré à l'utilisation d'électricité (du moins en France) demande beaucoup de ressources minérales.
- Un autre secteur d'impact moins important mais non négligeable, notamment sur l'usage des sols, sont les pauses café. Il faudrait cependant mettre cet impact en perspective avec l'impact de mon alimentation quotidienne, pour voir s'il est significatif.
- Les résultats d'ACV ont généralement assez peu de sens seuls, et sont en l'occurrence trop lacunaires pour être fiables.
- L'impact du stage seul semble malgré tout très minime : d'après cette étude taillée à la hache, j'ai davantage pollué pendant 6 mois d'utilisation des services publics qu'avec mon activité professionnelle [46].

7.2 Structure d'accueil

7.2.1 INRIA

L'INRIA est un institut de recherche public sur les sciences et technologies numériques. Il a donc, de par sa nature, des enjeux politiques, économique et sociaux à respecter. S'y retrouve également, bien entendu, une proximité forte avec la technologie et son usage.

Dans ce cadre, un penchant pour une approche environnementaliste pourrait sembler intuitivement prendre le pas sur une approche par l'écologie politique. Du moins en première approximation, puisqu'en réalité le bilan est davantage nuancé.

Comme premier point d'approche, plaçons le contexte : les objectifs de l'institut vis-à-vis de l'Etat [47]. 4 priorités ont été mises en avant pour l'INRIA :

- Maintenir l'excellence scientifique, notamment en encourageant la mise en place de nouveaux projets et partenariats, entre autre transdisciplinaires.
- Renforcer son impact économique, notamment via la création de startups.
- Construire une organisation sereine et efficace, notamment en essayant de lutter contre les discriminations au sein de l'institut.

- Renforcer son apport aux politiques publiques, par exemple en se rapprochant du milieu scolaire, mais également en se rapprochant de la sphère sécurité défense de l'Etat ¹⁶ et en participant à la transformation numérique de l'action publique.

Plusieurs dynamiques apparaissent donc en terme de société et d'environnement. Constatons d'abord que nous parlons d'un institut au service de la politique économique de l'Etat : ce n'est pas le lieu pour remettre en question cette dernière. Ceci dit, les employé·e·s de l'INRIA sont très libres de leur recherche et de leurs actions. Cet élément, ajouté à la volonté de l'institut d'encourager la prise de risque scientifique, offrent un contexte permettant à des équipes de recherche comme STEEP d'émerger, lesquelles peuvent se permettre de travailler sur de l'écologie et de proposer, à l'aide de la recherche scientifique, une critique constructive des problématiques sociétales actuelles.

Maintenant, certains éléments me posent question. Peut-on demander à un institut d'être propulseur de la croissance économique d'une part, et d'aider la société à rester dans les limites planétaires d'autre part [5] ? Peut-on attendre d'un institut dont l'une des priorités est avant tout l'excellence de ne pas reproduire des schémas de domination ? Peut-on avec certitude affirmer que le travail d'un institut poussant à prendre toujours plus de risques scientifiques, et encourageant les partenariats avec l'armée ¹⁷, soit toujours utilisé à des fins éthiques [48] ?

Quand l'Etat encourage à sa numérisation, a-t-il pensé à l'illectronisme [49] ? Quand le centre de Montbonnot est encouragé à faire de l'IOT (Internet Of Things), a-t-il pensé à l'empreinte écologique des objets connectés [16] ¹⁸ ?

Je n'ai pas toutes les réponses à ces questions, en tout cas, mon sentiment est qu'il y a, malgré le cadre, une forte volonté d'une part importante des employé·e·s de l'institut de s'intéresser aux questions écologiques. Par exemple, une critique du manque de prise en compte des impacts environnementaux de l'institut avait été émise lors de la parution de la première version du Contrat d'Objectifs et de Performance (COP) [47]. Cette critique avait mené à des modifications du-dit contrat. Une part importante du personnel se trouve également dans le collectif Ecoinfo, une instance mise en place par le CNRS (Centre National de la Recherche Scientifique) pour travailler sur la question de l'informatique écoresponsable.

Un document a également été rédigé en 2019 par un groupe de travail de l'INRIA appelé *MakeSEnS*, pour proposer un ensemble d'actions à mettre en place au niveau de l'institut sur les questions environnementales [51]. Ce document a permis de lancer une dynamique de mise en place d'un ensemble d'actions et de postes, notamment :

- Un poste d'adjoint·e du directeur ou de la directrice scientifique chargé·e des recherches sur l'environnement, au sein de la direction scientifique de l'INRIA : cette personne soutient les équipes interdisciplinaires sur l'écologie (à l'instar de STEEP). Elle aide par exemple au lien entre l'INRIA et l'INRAE (Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement), à proposer des thèses, à répondre à des appels à projet, etc.
- Un peu plus tard, a été mis en place un poste de chargé·e de mission environnementale au niveau de l'administration, dont le travail est de réfléchir au bilan environnemental de

16. J'ai d'ailleurs moi même assisté lors de mon passage à l'INRIA à une conférence des services français de renseignement ayant pour but de renforcer les partenariats.

17. Ce n'est bien entendu pas si simple, il y a une question de rapport de force entre pays qui entre en jeu ici.

18. En réalité, il existe un livre blanc des objets connectés [50], et ce dernier évoque vaguement (p.26) leurs impacts environnementaux, en disant que "À l'échelle mondiale, le déploiement de milliards d'appareils IoT consommera d'énormes quantités d'énergie et de ressources". L'auteur en conclut cependant que "L'IoT peut indéniablement être utile pour lutter contre le changement climatique."

l'institut, par exemple en analysant les missions des équipes avec un bilan carbone estimé, en mettant en place des politiques d'incitation à la réduction de ce bilan, etc.

- Une commission locale de développement durable, chargée de réduire l'impact de la vie au sein de l'institut, par exemple en organisant le retrait de la vaisselle jetable dans les cafétérias, en mettant en place des plans de déplacement, etc.
- Un groupe de travail sur l'équipement informatique, qui s'intéresse à conserver le plus longtemps possible les ordinateurs du centre, de pousser à l'achat de garanties, etc.

Parmi les autres actions effectuées par l'INRIA en ce sens, il est possible de noter la rédaction récente d'un livre blanc sur l'agriculture et le numérique [52] en partenariat avec l'INRAE. Il est également possible de noter qu'au niveau du centre de Montbonnot où je travaille (et probablement au niveau d'autres centres également), a été récemment effectué un bilan carbone.

En conclusion, le personnel de l'INRIA est en partie proche des questions écologiques et sociales, et la souplesse de l'institution permet à chacun et chacune d'expérimenter et d'orienter ses recherches dans cette direction. Une mobilisation existe réellement. Ceci dit, parmi le personnel sensibilisé, se trouvent des profils très variés, allant des écomodernistes [53] aux convivialistes [54]. Certains et certaines sont très proches d'une approche "tech for good"¹⁹, et d'autres très technocritiques, le tout se déployant dans un contexte encourageant davantage à l'environnementalisme qu'à l'écologie politique.

7.2.2 STEEP

STEEP est un exemple d'équipe projet interdisciplinaire à l'INRIA, ayant dévié des thématiques principales de l'institut (l'informatique et la robotique), pour s'orienter vers de tout autres champs disciplinaires : l'économie biophysique et l'écologie politique. L'importance de l'usage de la modélisation les fait cependant garder un lien avec le reste du centre.

Cette équipe a été fondée par Emmanuel PRADOS et Pierre-Yves LONGARETTI il y a une douzaine d'années, suite à un besoin fort des deux protagonistes de prendre du recul sur leur champs disciplinaires, pour se consacrer à une problématique bien plus urgente : les questions socioécologiques et les risques systémiques liés au fonctionnement de notre société actuelle.

Ce besoin, ressenti par les deux chercheurs, est fortement partagé par de nombreuses personnes issues des milieux intellectuels, que ça soit de la recherche ou de l'ingénierie par exemple. Le fait que l'équipe ait énormément grandi en un ou deux ans, passant d'une demi-douzaine à plus d'une vingtaine de personnes en est une preuve.

Si l'on est en droit de critiquer l'efficacité d'une institution comme STEEP dans l'urgence écologique actuelle, de par la lenteur qu'induit la recherche scientifique par exemple, et de par son potentiel de détournement par les structures politiques ; on ne peut cependant pas douter de la sincérité, de la prise de recul et de la capacité d'esprit critique dont font preuve ses membres, ni même de la ligne directrice des travaux qui en découle.

STEEP fait de l'écologie, et elle le fait bien.

Ses membres mettent en place de nombreuses initiatives visant à travailler l'écologie et la systémique à différentes échelles des milieux universitaires (et plus). Pour n'en citer que quelques unes, la mise en place des conférences *Comprendre et Agir*, l'université de printemps et la communauté *Antropocene Facts*, ou encore le symposium *Archipel*.

Autre élément notable, il y a une véritable volonté au sein de l'équipe de ne pas rester avec des travaux hors sol : leur objectif est de développer des outils participatifs, pour permettre aux

19. Tech For Good est une communauté cherchant à appliquer les nouvelles technologies à des objectifs jugés nobles, dans l'idée que la technologie peut aider à améliorer la société, à condition de bien l'orienter.

citoyens et citoyennes de prendre des décisions collectives pour l'avenir de la société. Plusieurs postes sont dédiés à cette problématique au sein de l'équipe : un postdoctorat est en cours, ainsi qu'un stage, et une thèse est à venir l'année prochaine, pour ne citer qu'eux.

Les membres de l'équipe sont en continuel développement de leur esprit critique et de leur prise de recul vis à vis des problématiques écologiques, et plus largement sociétales. Ces personnes mettent leurs compétences en oeuvre pour faire avancer la société dans une direction qui leur semble plus juste et plus résiliente.

7.3 Impact du projet

Essayons de procéder à l'analyse des impacts du projet de façon méthodologique. Commençons par les impacts les plus directs, et avançons vers les plus indirects.

À l'ordre 1, l'algorithme lié au projet est légèrement lourd en calcul, il consomme pour une utilisation standard autour de $25J$. En comparaison, une utilisation de la commande `echo` demande $0.0054J$. Ceci dit, un programme nécessitant de nombreux calculs, de nombreux appels au `kernel`, ou même possédant une interface graphique consommera davantage. Par exemple, l'ouverture (et uniquement l'ouverture, pas l'utilisation) de l'application de bureau Thunderbird consomme environ $47J$ dans le même contexte. Finalement, étant donné la nature du programme du PFE (lignes de commandes, avec un unique calcul important), celui-ci consomme effectivement une quantité d'énergie non négligeable, mais s'il est mis en comparaison avec à peu près n'importe quel autre type de programme, il ne consomme en réalité que peu²⁰. La majorité du calcul semble venir de l'appel aux backup `SQLite` de `Brightway2`. En revanche, si l'on regarde la consommation énergétique de la phase de setup du programme, la consommation énergétique de cette dernière est réellement importante, autour de $5473J$. Mais celle-ci n'est effectuée qu'une seule fois à chaque déploiement.

À l'ordre 2, l'objectif du programme est in fine d'être déployé sur un serveur, avec une interface web. Cette phase demandera, au niveau de l'usage, bien plus d'énergie que le simple programme local en lignes de commandes développé actuellement. Il sera alors intéressant de faire un diagnostic à l'aide d'un outil tel qu'Écoindex. Ceci dit, avoir un déploiement en ligne permettra de limiter le nombre de fois où la phase de setup est nécessaire (une seule fois contre une fois par utilisateur-riche auparavant).

À l'ordre 3, on remarque que le programme est destiné à une population de niche très spécifique. Il ne sera de toute façon que très peu utilisé, et uniquement par des personnes du milieu de la recherche (a priori). Son impact global en terme de consommation énergétique semble donc complètement négligeable. Aucun impact sociétal à ce niveau n'est non plus à noter : les personnes utilisant le programme sont des personnes averties, il n'est pas question ici d'illectronisme, ou de création de dépendance à un algorithme.

À l'ordre 4, la réflexion devient davantage intéressante. Ce programme va principalement servir à fournir des données à la modélisation `MATER`. La question qui se pose alors est : ces données sont-elles pertinentes ? Est-ce qu'utiliser des données d'ACV dans une modélisation fonctionnant complètement différemment ne risque pas de complètement fausser les résultats ? Auquel cas la modélisation n'aurait plus vraiment de sens, surtout si plusieurs jeux de données sont utilisés et non comparables. Or cette modélisation va potentiellement être réutilisée pour faire de la prospective, notamment dans l'objectif de faire des choix politiques de mise en place d'une transition. Il peut y avoir des conséquences écologiques et sociales fortes.

²⁰. Les mesures ont été effectuées sur les package 0 des `RAPL` (Running Average Power Limit) de mon ordinateur, à l'aide de l'outil `likwid-perfctr`.

À l'ordre 5, la question se pose de la pertinence de la modélisation énergétique en général. En effet, la modélisation a ses limites [55], et plus spécifiquement la modélisation prospective peut être mal comprise par des décideur·ses qui l'analyseraient comme de la prévision [56]. Il faut bien avoir conscience du cadre et des limites d'un modèle pour ne pas lui faire dire n'importe quoi, or ce recul demande généralement d'avoir suivi un cursus scientifique important, particulièrement pour des modélisations techniques avec de nombreuses données comme MATER. Ceci dit, certains éléments de participation peuvent être mis en place pour sensibiliser à ces questions, à l'instar de ce qu'essaye d'effectuer l'équipe STEEP.

Enfin, à l'ordre 6, se pose la question des résultats du modèle et de leur utilisation en politique. Ce point est intimement lié au point précédent. En effet, même très bien construit, un modèle ne donne que la vision d'un aspect spécifique de la réalité. Il est impossible de prévoir à l'avance ce que le modèle terminé donnera comme résultats. Prenons quelques exemples, et plaçons les dans un tableau qualitatif :

Résultat du modèle	Conclusion envisageable pro-transition	Conclusion envisageable anti-transition
Une transition est possible et ne pose aucun problème.	Faire la transition, sans questionner le modèle économique néolibéral.	Être dans le déni et continuer tel quel.
Une transition énergétique est possible mais nécessitera l'ouverture de nombreuses mines ayant des conséquences très importantes sur la santé humaine.	Faire la transition, sans regarder les conséquences humaines.	Justifier la non-transition par les conséquences soulevées par le modèle.
Une transition n'est pas envisageable telle quelle.	Éventuellement requestionner l'usage de l'énergie.	Justifier la non-transition par le modèle.

TABLE 5 – Quelques exemples de conséquences des résultats d'une modélisation prospective de transition énergétique

Comme le tableau 5 le montre, dans une approche d'écologie politique, il peut être risqué d'utiliser un modèle prospectif comme celui de MATER si la croissance économique n'est pas remise en question. Or, lorsque les sphères politiques demandent à ce type de modèle de donner des résultats prospectifs, une condition inévitablement posée est un certain taux de croissance (en général 2%), comme il en a notamment été le cas pour le modèle *World 3* [57]. Si l'on considère que la croissance économique, ou a minima nos besoins en énergie, doivent être remis en question, le seul cas de figure qui semble aboutir à des conclusions de ce type est celui où le modèle déclare qu'il est impossible d'effectuer une transition dans un modèle de croissance économique (à l'instar du modèle *World 3* [5]), et que les personnes qui analysent ce résultat soient dans une perspective écologiste.

Il serait alors légitime de se demander : pourquoi ne pas vouloir de croissance si le modèle dit qu'il est possible d'en avoir ? Et effectivement, si le modèle déclare qu'une transition est possible avec croissance, dans ce cadre, il n'y a pas de remise en question du modèle économique néolibéral. Mais dans ce cadre uniquement, puisque le modèle a ses limites, ne regarde qu'une dimension du problème, et vouloir croire qu'une simple transition énergétique suffirait à grader

notre modèle de société signifie ne pas avoir une vision systémique du problème. Il existe en effet des effets rebonds, des conséquences écologiques et sociales fortes qui ne sont pas comprises dans cette problématique, etc.

Finalement, la question qui se pose est : est-il pertinent de parler de transition énergétique ? Historiquement, de telles transitions ne semblent jamais avoir eu lieu à l'échelle globale. Pire, elles semblent avoir servi à cacher la problématique de la pollution au profit de celle de l'épuisement des ressources [11]. Cependant, à l'échelle locale, il semble que des transitions énergétiques aient réussi à avoir lieu [58], ce qui pose la question des imports et des exports, et notamment de la dépendance des pays les plus développés envers les pays les moins développés [59]. Dans tous les cas, la transition énergétique n'est qu'une partie des problématiques d'émission de CO_2 , qui n'est elle-même qu'une partie des problèmes environnementaux. Il s'agit pourtant quasiment du seul aspect de la question écologique étudiée par le débat public, d'où le risque de conclusions décalées présentées dans le tableau 5.

Synthèse et conclusion

Notre époque est marquée par une urgence sociale et environnementale. Dans ce cadre, la recherche universitaire se mobilise de plus en plus pour aider le grand public à trouver des solutions (section 1.1).

L'une des méthodes les plus utilisées par les scientifiques pour quantifier les impacts environnementaux d'un système donné est l'Analyse de Cycle de Vie (ACV). Cette méthodologie se construit selon une logique entrée-sortie, en s'appuyant sur des bases de données gigantesques qui implémentent et chiffrent des graphes représentant des chaînes de production (section 2.2).

Une autre méthode de quantification des impacts environnementaux consiste en la modélisation prospective. C'est le cas du projet MATER, proposé par une équipe du laboratoire des sciences de la terre de Grenoble, ISTerre (section 2.3).

C'est dans ce contexte que s'est construit le stage, ayant pour but de développer un outil servant à grouper des sous-graphes de production des bases de données d'ACV. Ces groupements serviront ensuite à récupérer des données pour la modélisation MATER (section 2.4).

Une première phase importante du stage aura été une longue période d'échanges avec les comanditaire-rices, afin d'identifier clairement les besoins de chacun et chacune et d'en tirer un objectif clair et atteignable pour le stage (annexe D).

Deux grandes phases ont ensuite marqué le déroulement du stage : le développement de l'algorithme de groupement d'une part, et la gestion du système de test d'autre part. L'algorithme de groupement se sera finalement soldé par une logique entrée-sortie similaire à celle de l'ACV (section 4). Le système de test aura quant à lui été construit sur une logique d'indicateurs et de seuils (section 5), permettant d'assurer une certaine flexibilité à ses résultats, lesquels sont instables en raison du mauvais conditionnement de l'algèbre de l'ACV (section 5.2).

De nombreux développements restent aujourd'hui à effectuer, pouvant faire l'objet d'un futur stage (section 6). En effet, contrairement à ce qui était prévu dans le planning prévisionnel, la gestion du système de test aura pris énormément de temps, ne laissant pas assez de place pour terminer tous les objectifs de stage fixés initialement (annexe D).

Finalement, ce stage m'aura apporté beaucoup à titre personnel, me tirant fortement vers le haut au niveau, entre autre, de ma culture autour de l'écologie. Il aura également été formateur sur la période d'identification des besoins, et de co-construction des objectifs du stage. Techniquement j'aurai découvert un certain nombre de notions, par exemple celle de conditionnement matriciel. Mais ce stage aura été avant tout riche en rencontres et en perspectives, puisqu'il y a de grandes chances que grâce à lui, je continue en thèse.

Remerciements

Je dois beaucoup à de nombreuses personnes pour la réussite de ce stage.

Je tiens tout d'abord à remercier mes tuteurs de stage, Guillaume MANDIL et Bernard TOURANCHEAU, ainsi que mon tuteur d'école, Sylvain BOUVERET. C'est avant tout grâce à leur encadrement que le stage s'est déroulé sans encombres.

Travailler avec Jeanne BARRIOL aura été un plaisir, un grand merci à elle pour sa bonne humeur et son professionnalisme.

Je dois une fière chandelle à un bon nombre de membres de STEEP pour m'avoir accompagné dans mon travail. Un grand merci à Peter STRUM pour m'avoir initié au conditionnement matriciel, et pour m'avoir aidé à identifier la politique environnementale de l'INRIA. Merci également à Jean-Yves COURTONNE pour m'avoir accompagné durant le stage, et tout particulièrement pour avoir effectué la démonstration de fonctionnement de mon code. Je tiens également à remercier Vincent JOST, qui m'a aidé sur des problématiques de jeux de test, et Mathieu MANGEOT qui m'a accompagné et m'a mis à disposition un serveur, dont malheureusement je n'aurai pas eu le temps de me servir. Une mention spéciale à Léon FAUSTE pour ses coups de main réguliers, à Yazid CHARKANI et à Antonin BERTHE pour les papiers qu'ils m'ont transmis.

Merci énormément également à Marie-Anne DAUPHIN pour s'être occupée de toutes les démarches administrative

Merci à l'équipe d'ISTerre, et tout particulièrement à François VERZIER qui encadrait le stage de Jeanne, et donc, dans une certaine mesure, le mien également.

Et surtout, un énorme merci à toutes les personnes de l'équipe avec qui j'ai pu interagir, lesquelles ont chacune contribué à me faire me sentir bien, et à me tirer vers le haut intellectuellement. On y compte, en plus des personnes déjà mentionnées, Emmanuel P, Pierre-Yves, Niels, Sophie, Mathilde B, Alexandre B, Grégoire, Louis, Quentin, Mathilde dP, Olivier, François-Rémi, Emmanuel K, Léa, Arianne, Enzo et Alexandre P.

J'espère avoir dans le futur l'occasion de travailler à nouveau avec toutes ces personnes, tant leur compagnie aura marqué un tournant fort dans ma vie.

Enfin, un tout dernier merci à mes relecteurs et relectrices, qui contribuent fortement à la montée en qualité de ce rapport.

Références bibliographiques

- [1] C. Chavagneux, “Les quatre crises françaises,” *L’Economie politique*, vol. 58, no. 2, pp. 5–5, 2013.
- [2] B. Sananes, L. Bedeau, and V. Thomas, “Les « gilets jaunes » : la partie émergée de la crise sociale française ?,” 2019.
- [3] IPCC, *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.
- [4] B. Manin, *Principes du gouvernement représentatif*. Calmann-Lévy, 2012.
- [5] D. H. Meadows, D. H. Meadows, J. Randers, and W. W. Behrens III, “The limits to growth : a report to the club of rome (1972),” *Google Scholar*, vol. 91, 1972.
- [6] K. Raworth, *Doughnut economics : seven ways to think like a 21st-century economist*. Chelsea Green Publishing, 2017.
- [7] S. Wuhl, “A propos des fondements de la démocratie participative,” 2009.
- [8] R. Grandjean, M. Lepetit, and L. Morel, “Scénarios énergie-climat : Évaluation et mode d’emploi,” tech. rep., The Shift Project and Afep, 2019.
- [9] “La transition énergétique au cœur d’une transition sociétale,” tech. rep., négaWatt, 2021.
- [10] B. Mayer, “Enjeux et résultats de la cop21,” *Revue juridique de l’environnement*, vol. 41, no. 1, pp. 13–17, 2016.
- [11] J.-B. Fressoz, “Pour une histoire désorientée de l’énergie,” in *25èmes Journées Scientifiques de l’Environnement-L’économie verte en question*, no. 04, 2014.
- [12] L. Gigandon, “Automating supply chain analysis using life cycle assessment databases,” 2018.
- [13] A. Borthomieu, “Material and energy flows of the french economy,” 2020.
- [14] W. Leontief, “Environmental repercussions and the economic structure : an input-output approach,” *The review of economics and statistics*, pp. 262–271, 1970.
- [15] B. Weidema, H. Wenzel, C. Petersen, and K. Hansen, “The product, functional unit and reference flows in lca,” *Environmental News*, vol. 70, pp. 1–46, 2004.
- [16] F. Bordage, “Empreinte environnementale du numérique mondial,” *Paris, greenit. fr*, p. 9, 2019.
- [17] G. Rebitzer, T. Ekvall, R. Frischknecht, D. Hunkeler, G. Norris, T. Rydberg, W.-P. Schmidt, S. Suh, B. P. Weidema, and D. W. Pennington, “Life cycle assessment : Part 1 : Framework, goal and scope definition, inventory analysis, and applications,” *Environment international*, vol. 30, no. 5, pp. 701–720, 2004.
- [18] S. Nakamura and Y. Kondo, *Waste input-output analysis : concepts and application to industrial ecology*. Springer, 2009.
- [19] V. Colomb, S. Ait-Amar, C. Basset-Mens, A. Gac, G. Gaillard, P. Koch, J. Mousset, T. Salou, A. Tailleur, and H. M. Van Der Werf, “Agribalyse®[®], the french lci database for agricultural products : high quality data for producers and environmental labelling,” 2015.
- [20] R. Frischknecht, N. Jungbluth, H.-J. Althaus, G. Doka, R. Dones, T. Heck, S. Hellweg, R. Hischier, T. Nemecek, G. Rebitzer, *et al.*, “The ecoinvent database : overview and methodological framework (7 pp),” *The international journal of life cycle assessment*, vol. 10, no. 1, pp. 3–9, 2005.

- [21] S. Joshi, “Product environmental life-cycle assessment using input-output techniques,” *Journal of industrial ecology*, vol. 3, no. 2-3, pp. 95–120, 1999.
- [22] R. Heijungs, “Is mainstream lca linear ?,” *The International Journal of Life Cycle Assessment*, vol. 25, pp. 1–11, 10 2020.
- [23] O. Vidal, “Fiche de stage : Implémentation d’un module cycle de vie sur une modélisation dynamique pour la transition énergétique,” 2022.
- [24] F. Verzier, “Modélisation dynamique multi-régionale de la demande en énergie, matériaux et infrastructures,” 2021.
- [25] R. Sacchi, T. Terlouw, K. Siala, A. Dirnaichner, C. Bauer, B. Cox, C. Mutel, V. Daioglou, and G. Luderer, “Prospective environmental impact assessment (premise) : A streamlined approach to producing databases for prospective life cycle assessment using integrated assessment models,” *Renewable and Sustainable Energy Reviews*, vol. 160, p. 112311, 2022.
- [26] E. A. Parson, A. Fisher-Vanden, and Karen, “Integrated assessment models of global climate change,” *Annual Review of Energy and the Environment*, vol. 22, no. 1, pp. 589–628, 1997.
- [27] C. Mutel, *Wurst documentation*. <https://wurst.readthedocs.io/>.
- [28] C. Mutel, “Brightway : an open source framework for life cycle assessment,” *Journal of Open Source Software*, vol. 2, no. 12, p. 236, 2017.
- [29] C. Mutel, *Brightway2 documentation*. <https://2.docs.brightway.dev/>.
- [30] R. Speck, S. Selke, R. Auras, and J. Fitzsimmons, “Life cycle assessment software : Selection can impact results,” *Journal of Industrial Ecology*, vol. 20, no. 1, pp. 18–28, 2016.
- [31] I. T. Herrmann and A. Moltesen, “Does it matter which life cycle assessment (lca) tool you choose?—a comparative assessment of simapro and gabi,” *Journal of Cleaner Production*, vol. 86, pp. 163–169, 2015.
- [32] A. Citroth, “Ict for environment in life cycle applications openlca—a new open source software for life cycle assessment,” *The international journal of life cycle assessment*, vol. 12, no. 4, pp. 209–210, 2007.
- [33] B. Steubing, D. de Koning, A. Haas, and C. L. Mutel, “The activity browser—an open source lca software building on top of the brightway framework,” *Software Impacts*, vol. 3, p. 100012, 2020.
- [34] C. Mutel, *Brightway2 key concepts*. <https://2.docs.brightway.dev/intro.html>.
- [35] J. Deacon, “Model-view-controller (mvc) architecture,” *Online* [Citado em : 10 de março de 2006.] <http://www.jdl.co.uk/briefings/MVC.pdf>, 2009.
- [36] N. J. Higham, *Accuracy and stability of numerical algorithms*. SIAM, 2002.
- [37] R. Takapoui and H. Javadi, “Preconditioning via diagonal scaling,” *arXiv preprint arXiv :1610.03871*, 2016.
- [38] R. L. Harrison, “Introduction to monte carlo simulation,” in *AIP conference proceedings*, vol. 1204, pp. 17–21, American Institute of Physics, 2010.
- [39] G. Carbou, “En finir avec la confusion entre écologie et environnementalisme,” *Libération*, Jun 2019.
- [40] I. Pypaert Perrin, M. Hosselet-Herbignat, B. Tardieu, E. Gaillard, H. Totin, M. Pochon, G. Wallenborn, S. Swaton, G. Malfait, and al, *Justice, climat : même combat !* Éditions Quart Monde, Feb 2019.

- [41] S. Persico, *Un clivage, des enjeux : une étude comparée de la réaction des grands partis de gouvernement face à l'écologie*. Theses, Institut d'études politiques de paris - Sciences Po, Nov. 2014.
- [42] M. Bookchin and B. Weigel, *Qu'est-ce que l'écologie sociale ?* Atelier de création libertaire, 2012.
- [43] P. Descola, *Par-delà nature et culture*, vol. 1. Gallimard Paris, 2005.
- [44] Etudiant·e-s, "Appel à désertier - remise des diplômes agroparistech 2022," 2022.
- [45] C. Alexandre, A. Gérard, M. Goedkoop, and T. C. Ponsioen, *Indicateurs d'impact environnementaux en ACV : état de l'art, retour d'expériences et recommandations*. PhD thesis, SCORE LCA, 2014.
- [46] "Empreinte carbone des services publics," *Bon Pote*, feb 2021.
- [47] Ministère de l'économie et des finances, Ministère de l'enseignement supérieur de la recherche et de l'innovation, "Contrat d'objectifs et de performance 2019-2023," 2019.
- [48] A. Grothendieck, "Allons nous continuer la recherche scientifique ?," 1972.
- [49] P. Brotcorne, "Digitalisation des services publics : la face sombre," *Revue Quart Monde*, pp. 4–7, april 2018.
- [50] E. Baccelli, "Internet des objets : Défis sociétaux et domaines de recherche scientifique pour l'internet des objets (iot)," 2021.
- [51] F. Berthoud, P. Guitton, L. Lefèvre, S. Quinton, A. Rousseau, J. Sainte-Marie, C. Serrano, J.-B. Stefani, P. Sturm, and E. Tannier, "Sciences, environnements et sociétés," 2019.
- [52] V. B. Maurel, F. Garcia, L. Brossard, N. Mitton, and A. Termier, "Agriculture et numérique- livre blanc inria-inrae," 2022.
- [53] "An ecomodernist manifesto." <http://www.ecomodernism.org/manifesto-english>.
- [54] I. Illich, L. Giard, and V. Bardet, *La convivialité*. Editions du Seuil Paris, 1973.
- [55] M. Buchanan, *The limits of a model*. PhD thesis, Nature Publishing Group, 2020.
- [56] M. Godet *et al.*, "Prospective : sept idées-clés et une méthode," tech. rep., 1983.
- [57] B. De Vries and P.-Y. Longaretti, "Lecture on models – where we are, lecture during antropocene facts spring scool," 2022.
- [58] B. K. Sovacool, "How long will it take? conceptualizing the temporal dynamics of energy transitions," *Energy Research & Social Science*, vol. 13, pp. 202–215, 2016.
- [59] M. Manfroni, R. Velasco-Fernández, L. Pérez-Sánchez, S. G. Bukkens, and M. Giampietro, "The profile of time allocation in the metabolic pattern of society : An internal biophysical limit to economic growth," *Ecological Economics*, vol. 190, pp. 107–183, 2021.

A Démonstration formelle de l'algorithme de groupement

Cette démonstration est à l'ordre 3 (3 activités dans la base de données), avec un groupement d'ordre 2 (2 activités à grouper) afin de faciliter la compréhension et le raisonnement. Elle s'étend cependant facilement à l'ordre n .

Pour des facilités de notation et par respect pour les conventions, les activités sont ici nommées "Secteurs" et sont notés $S_i, i \in \{1, 2, 3\}$. En effet, dans la logique de Leontief, il s'agit bien des secteurs qui sont les blocs unitaires entrée-sortie, car le raisonnement est macroéconomique. L'ACV, elle, utilise davantage les termes de "Processus" ou d'"Activités".

Cette démonstration a été effectuée par Jean-Yves COURTONNE.

a Définition des termes

$$A = \begin{matrix} & S_1 & S_2 & S_3 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} & \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \end{matrix}$$

FIGURE A.1 – Matrice des coefficients : la colonne j donne la recette pour produire une unité de secteur j .

$$L = \begin{matrix} & S_1 & S_2 & S_3 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} & \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \end{matrix}$$

FIGURE A.2 – Matrice de Leontief : la colonne j indique tout ce qu'il faut produire dans la chaîne d'approvisionnement pour pouvoir consommer une unité de j .

$$L = (I - A)^{-1} = I + A + A^2 + A^3 + \dots + A^n + \dots$$

$$L_{ij} = \delta_{ij} + A_{ij} + \sum_k A_{ik}A_{kj} + \sum_{k,l} A_{ik}A_{kl}A_{lj} + \dots$$

$$\hat{A} = \begin{matrix} & S_2 & S_3 \\ \begin{matrix} S_2 \\ S_3 \end{matrix} & \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} \end{matrix}$$

FIGURE A.3 – Matrice réduite des coefficients.

$$\hat{L} = \begin{matrix} & S_2 & S_3 \\ S_2 & \hat{L}_{22} & \hat{L}_{23} \\ S_3 & \hat{L}_{32} & \hat{L}_{33} \end{matrix}$$

FIGURE A.4 – Matrice réduite de Leontief.

Notons que $L_{ij} \neq \hat{L}_{ij}$.

Soit x , l'ICV d'une unité de S_3 . Il est donné par :

$$x = ICV(1 \text{ unite } S_3) = BLf_3$$

avec B la matrice des flux de biosphere (de taille nombre de flux de biosphere \times nombre de secteurs) et :

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, f_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, f_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

b Démonstration

$$A = \begin{pmatrix} a & b \\ c & \hat{A} \end{pmatrix} \quad \text{et} \quad I - A = \begin{pmatrix} I - a & -b \\ -c & I - \hat{A} \end{pmatrix}$$

Pour l'inversion d'une matrice M par blocs, on utilise la formule duale (si D est également inversible) :

$$M^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & D^{-1} \end{pmatrix} + \begin{pmatrix} I \\ -D^{-1}C \end{pmatrix} (A - BD^{-1}C)^{-1} (I \quad -BD^{-1})$$

Donc, en partant de cette formule duale :

$$L = \begin{pmatrix} 0 & 0 \\ 0 & \hat{L} \end{pmatrix} + \begin{pmatrix} I \\ \hat{L}c \end{pmatrix} (I - a - b\hat{L}c)^{-1} (I \quad b\hat{L})$$

Soit

$$L = \begin{pmatrix} L_\alpha & L_\beta \\ L_\gamma & L_\delta \end{pmatrix}$$

On voit que :

$$\begin{pmatrix} I \\ \hat{L}c \end{pmatrix} (I - a - b\hat{L}c)^{-1} (I \quad b\hat{L}) = \begin{pmatrix} L_\alpha & L_\beta \\ L_\gamma & L_\delta - \hat{L} \end{pmatrix}$$

puis :

$$\begin{pmatrix} I \\ \hat{L}c \end{pmatrix} (I - a - b\hat{L}c)^{-1} = \begin{pmatrix} L_\alpha \\ L_\gamma \end{pmatrix}$$

L'ICV limité à la technosphère correspond à la dernière (ici troisième, mais dans le cas général n-ième) colonne de L . On a donc :

$$Lf_n = \begin{pmatrix} 0 & 0 \\ 0 & \hat{L} \end{pmatrix} f_n + \begin{pmatrix} L_\alpha \\ L_\gamma \end{pmatrix} (I \quad b\hat{L}) f_n$$

Dans cette relation, le terme $\begin{pmatrix} 0 & 0 \\ 0 & \hat{L} \end{pmatrix} f_n$ représente les besoins intrinsèques du secteur groupé envers les technosphères des secteurs qui le constitue, en d'autres termes, ce terme est proportionnel à l'impact direct du secteur groupé.

$(I \ b\hat{L}) f_n$ est un vecteur colonne (le nombre de lignes correspond au nombre de secteurs hors secteurs groupés), qui représente les besoins directs du secteur groupé pour les secteurs non groupés. Par définition, il suffit de multiplier par les colonnes de la matrice L correspondant aux secteurs non groupés soit précisément $\begin{pmatrix} L_\alpha \\ L_\gamma \end{pmatrix}$ pour obtenir l'ICV technosphère correspondant (indice α = en secteurs non groupés, indice γ = en secteurs groupés).

Finalement, l'égalité ci-dessus s'interprète de la même façon que $L = I + LA$.

Pour l'ICV exprimé en biosphère, la démonstration est immédiate : il suffit de multiplier à gauche par la matrice B .

B Tableau des indicateurs de test

Vecteur	Indicateur statistique	Seuil(s)	Variable(s) config.ini	Moyenne itérative
Erreurs absolues	Moyenne			
Erreurs absolues	Médiane			
Erreurs absolues	1 ^{er} quartile			
Erreurs absolues	3 ^{ème} quartile			
Erreurs absolues	Variance			
Erreurs absolues	Maximum	W > 10 ⁻¹ E > 10 ²	ABSOLUTE_TOLERANCE_FACTOR ABSOLUTE_EXTREME_TOLERANCE_FACTOR	Géométrique
Erreurs absolues	Minimum	W > 0	ABSOLUTE_MIN_TOLERANCE_FACTOR	Géométrique
Erreurs absolues	<i>pvalue</i> du test du χ^2 de la loi de Benford	W < 0.9999	ABSOLUTE_BENFORD_TOLERANCE_FACTOR	Arithmétique
Erreurs relatives	Moyenne			
Erreurs relatives	Médiane			
Erreurs relatives	1 ^{er} quartile			
Erreurs relatives	3 ^{ème} quartile			
Erreurs relatives	Variance			
Erreurs relatives	Maximum	W > 10 ⁻³ E > 10 ⁻¹	RELATIVE_TOLERANCE_FACTOR RELATIVE_EXTREME_TOLERANCE_FACTOR	Géométrique
Erreurs relatives	Minimum			
Erreurs relatives	<i>pvalue</i> du test du χ^2 de la loi de Benford	W < 0.1	RELATIVE_BENFORD_TOLERANCE_FACTOR	Arithmétique
	Temps de compilation			
	Temps d'ICV de référence			
	Temps d'ICV après groupement			
	Singularités	E > 0.9	SINGULARITY_TOLERANCE_FACTOR	Arithmétique

TABLE B.1 – Table des indicateurs statistiques sur les jeux de tests

Légende

- W : Un warning est levé si le seuil est dépassé
- E : Une erreur est levée si le seuil est dépassé

C Comparatif de la répartition du premier chiffre significatif sur des vecteurs d'erreurs absolues et des vecteurs d'erreurs relatives

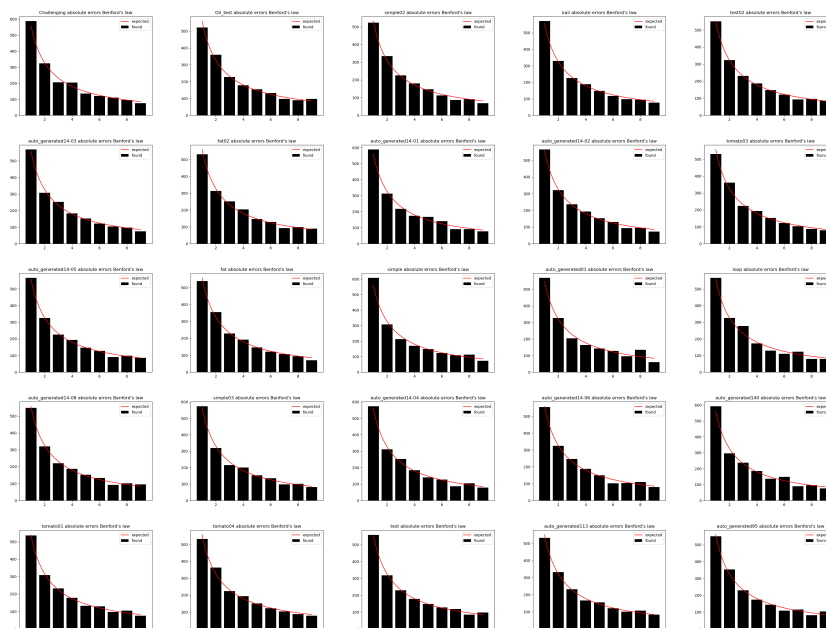


FIGURE C.1 – Graphes de répartition du premier chiffre significatif des erreurs absolues : jeu de 25 tests

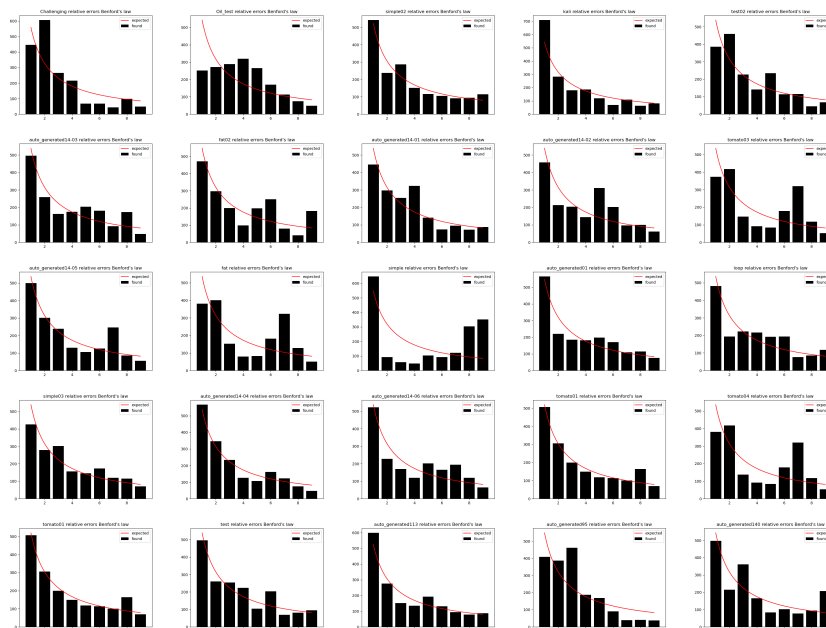


FIGURE C.2 – Graphes de répartition du premier chiffre significatif des erreurs relatives : jeu de 25 tests

D Organisation du travail

a Diagramme de Gantt

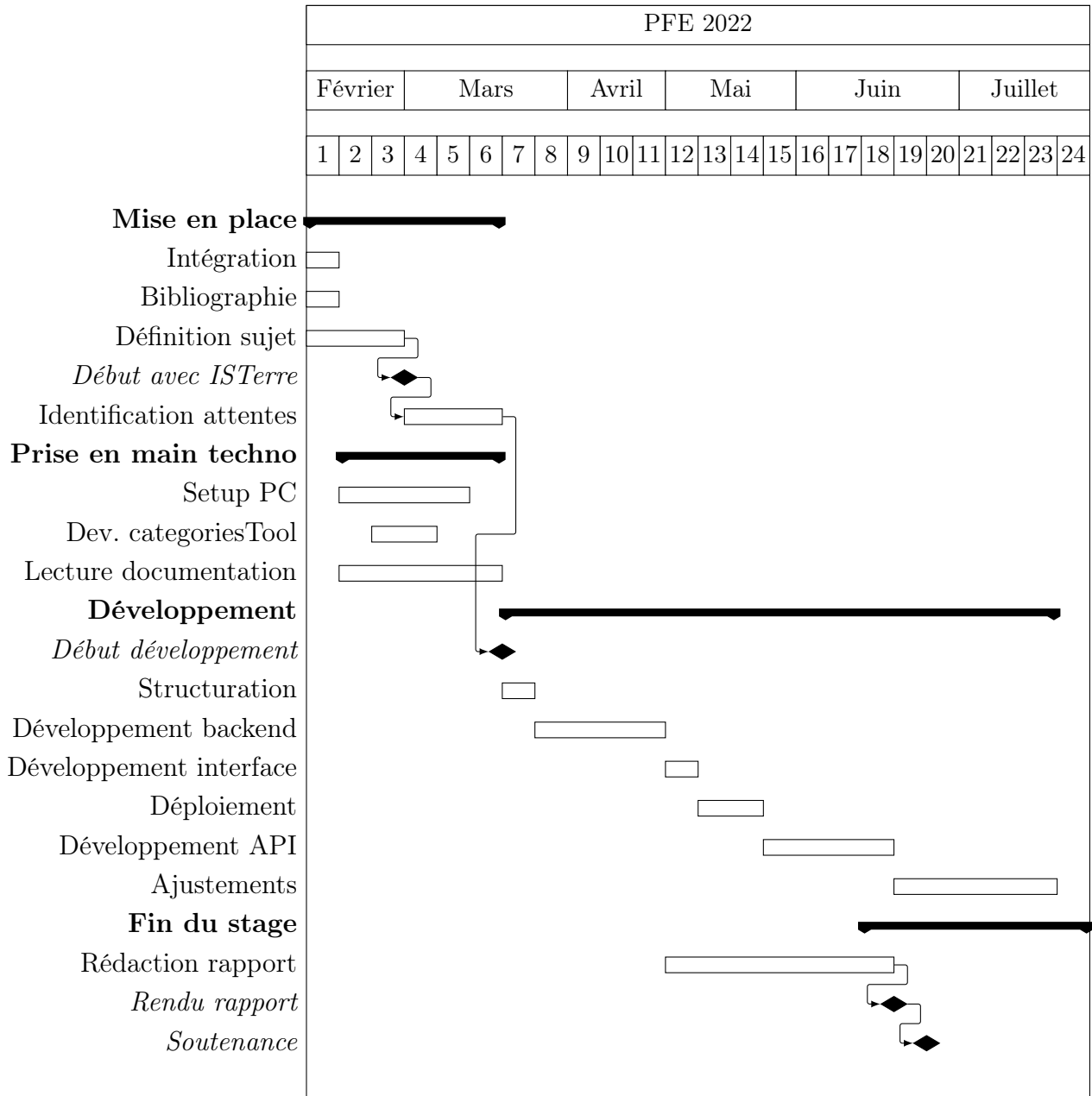


FIGURE D.1 – Diagramme de Gantt, en semaines

b Fonctionnement du projet

Comme présenté dans la section 1.3, le stage est au service du projet MATER, et plus spécifiquement de François VERZIER, un doctorant de l'équipe de MATER.

Le stage s'effectue en collaboration avec Jeanne BARRIOL, stagiaire du côté d'ISTERre qui a également pour mission la réconciliation des bases de données d'ACV avec la modélisation

MATER, mais dans une perspective théorique, à la différence de l'approche technique qui concerne ce stage.

Des réunions régulières avec François et Jeanne ont lieu pour échanger sur les attentes et les compétences des un·e·s et des autres. C'est suite à presque un mois d'échanges que le projet a réussi à se dessiner concrètement.

Beaucoup d'échanges ont également lieu avec les autres stagiaires de STEEP qui travaillent sur des problématiques similaires, notamment de réconciliation du modèle ACV avec celui de diagrammes de flux de matière.

Des points réguliers sont effectués avec les tuteurs de stage afin d'échanger des questions et d'assurer l'avancement du projet.

c Journée type

Une journée type de stage se structure globalement comme suit :

- **8h25** : Départ de l'appartement à vélo
- **9h10** : Arrivée à l'INRIA
- **9h10-12h10** : Travail
- **12h10-13h30** : Pause midi
- **13h30-17h30** : Travail
- **17h30** : Départ de l'INRIA
- **18h10** : Arrivée à l'appartement

Ce type de journée est à géométrie très variable : elle dépend chaque jour des personnes présentes au bureau, des réunions (notamment tous les mardi matin de **10h30** à **12h**, soit spécifiques à l'axe AST, soit de l'ensemble de l'équipe une fois par mois), des réunions à ISTerre (sur le campus), etc..

J'effectue des journées en distanciel environ une fois par semaine (généralement le vendredi car les bureaux sont plus ou moins vides ce jour-ci, or je n'ai pas les clés).

E Retour d'expérience sensible du pré-rapport

Après deux mois de stage, je suis très satisfait de travailler à STEEP. L'équipe est particulièrement agréable et me tire personnellement fortement vers le haut, autant dans l'évolution de mon esprit critique et de ma pensée systémique que dans mes réflexions personnelles sur une suite en thèse.

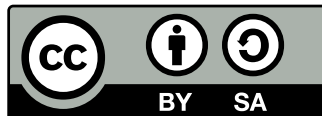
J'apprends également à échanger avec les personnes concernées afin d'en dégager une attente précise, et du point de vue technique à utiliser certaines librairies, à configurer une machine, à effectuer un déploiement via `Docker`, etc..

Du point de vue des problèmes, ils auront davantage été d'ordre technique : `Brightway2` étant assez peu mis à jour, j'ai été confronté à un certain nombre d'erreurs liées à des questions de compatibilité. Cela m'aura permis de me rendre utile sur cette question-là aux autres stagiaires qui avaient besoin de `Brightway2`.

En conclusion, le stage se passe bien, semble avancer à un rythme cohérent et me plaît.

Sauf mention contraire, cet ouvrage et son contenu texte sont sous licence “Attribution - Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0)”. Veuillez citer comme :

De Goër, Baptiste, “Dépendances technologiques : Développement d’un système de liaison entre le modèle d’Analyse de Cycle de Vie (ACV) et la modélisation MATER”, STEEP (2022) / CC BY-SA 4.0



Sauf mention contraire et à l’exception des logos d’institution, les illustrations de cet ouvrage sont sous licence “CC0 1.0 universel (CC0 1.0)”.

