Titre: Title:	Analyse de la propagation des incertitudes et données agrégées en analyse du cycle de vie : évaluation et prise en compte de différents niveaux de corrélations
Auteur: Author:	Yohan Marfoq
Date:	2015
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Marfoq, Y. (2015). Analyse de la propagation des incertitudes et données agrégées en analyse du cycle de vie : évaluation et prise en compte de différents niveaux de corrélations [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/1788/

UNIVERSITÉ DE MONTRÉAL

ANALYSE DE LA PROPAGATION DES INCERTITUDES ET DONNÉES AGRÉGÉES EN ANALYSE DU CYCLE DE VIE : ÉVALUATION ET PRISE EN COMPTE DE DIFFÉRENTS NIVEAUX DE CORRÉLATIONS

YOHAN MARFOQ
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)
JUIN 2015

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ANALYSE DE LA PROPAGATION DES INCERTITUDES ET DONNÉES AGRÉGÉES EN ANALYSE DU CYCLE DE VIE : ÉVALUATION ET PRISE EN COMPTE DE DIFFÉRENTS NIVEAUX DE CORRÉLATIONS

présenté par : MARFOQ Yohan
en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de :

M. PARTOVI NIA Vahid, Doctorat, président
M. MARGNI Manuele, Doctorat, membre et directeur de recherche
Mme BULLE Cécile, Ph. D., membre

DÉDICACE

À ma famille, mes amis
et à tous ceux qui protègent notre Terre...

REMERCIEMENTS

Je tiens en premier lieu à remercier mon directeur de maîtrise Manuele Margni, ainsi que Pascal Lesage qui m'ont soutenu de manière constante durant toute ma recherche et dont les conseils ont été d'une aide cruciale.

Je souhaite ensuite remercier l'intégralité des membres du CIRAIG dont la cohésion et l'entraide autour des valeurs communes du partage et du développement durable ont été d'une grande aide. Plus particulièrement un grand merci à mes camarades de bureau avec qui j'ai partagé la majeure partie de mon temps : Laure, Leyla, Stephanie, Viet et Vincent.

Ensuite, j'aimerais remercier ma famille qui malgré la distance a su rester proche de mon coeur et m'a toujours soutenu et conseillé dans mes choix.

Finalement, un grand merci à mes colocataires et tous mes amis sans qui ces deux premières années au Canada n'auraient jamais été aussi plaisantes.

RÉSUMÉ

Les enjeux environnementaux placent le développement durable au coeur des problématiques des entreprises industrielles. C'est dans le cadre des normes ISO 14040 et ISO 14044 que l'Analyse de Cycle de Vie (ACV) s'inscrit comme un outil permettant d'aider ces industries en les éclairant sur leurs interactions avec l'environnement.

L'étude s'inscrit dans l'utilisation des outils simplifiés ACV afin de compléter les résultats d'impact obtenus avec des informations sur leurs incertitudes, car elles représentent un élément clé pouvant altérer les conclusions.

L'objectif général de l'étude est de donner des informations d'incertitudes sur les résultats d'impacts d'une Analyse de Cycle de Vie, tout en considérant différents niveaux de corrélations intervenant dans la propagation de ces incertitudes au sein du cycle de vie du produit ou du service. Afin d'arriver à satisfaire cet objectif général, l'étude se divise en quatre objectifs spécifiques :

- L'évaluation de la sensibilité d'une hypothèse de corrélation intradonnées qui est généralement faite dans le calcul des incertitudes des impacts en ACV. Pour cela les informations d'incertitudes sont déterminées avec puis sans cette hypothèse, puis les différents résultats obtenus sont comparés.
- La prise en compte de corrélation (intrasystème) entre les différents impacts agrégés des processus utilisés dans les logiciels ACV simplifiés lors de l'évaluation des incertitudes des impacts d'un système complet. La part des covariances dans la variance totale du système est déterminée afin de savoir son importance.
- La prise en compte de la corrélation (intersystème) entre deux systèmes de produits agrégés lors de la comparaison de leurs impacts. Cela permet de nuancer la comparaison des impacts de deux systèmes de produits afin de savoir quelle est la probabilité que l'un soit meilleur que l'autre.
- La réalisation d'une preuve de concept avec une étude de cas proposé par un partenaire industriel. L'outil réalisé doit répondre aux attentes des industriels en terme de simplicité d'utilisation, de clarté des résultats et de rapidité d'exécution des calculs.

La méthodologie utilisée consiste à évaluer ces trois niveaux de corrélation lors de la propagation des incertitudes, puis de les considérer dans les résultats apportés par l'outil qui est réalisé.

L'étude réalisée a permis de savoir que la propagation des impacts n'est relativement pas sensible à l'hypothèse de corrélation intradonnées. Cependant, la corrélation intrasystème change considérablement les résultats des incertitudes des impacts d'un système de produit. De plus, l'outil développé permet de vérifier que les attentes des industriels en terme de simplicité, rapidité et clarté sont atteintes. Il est donc possible d'intégrer des informations d'incertitudes aux logiciels ACV simplifiés tout en considérant les différents niveaux de corrélations présentés dans l'étude.

ABSTRACT

The environmental stake chanllenges the requierments of the industrial buisnesses. Within the framework of ISO 14040 and ISO 14044 standard, the Life Cycle Assessment (LCA) is a tool helping industries to know how they interact with the environment.

The research frames in the use of simplified tools, the LCA in order to obtain the impact results with the informations on the uncertainty, which is a key when drawing conclusions.

The main purpose of the research is to set the uncertainties of the impact results of an LCA, considering the different correlation levels involved in the propagation of the uncertainties within the life cycle of a product or service. To satisfy this aim, the research is divided in four goals.

– The estimation of the sensitivity of an intradata correlation assumption is computed with the uncertainties impacts in LCA. To do this, the information concerning the uncertainties are determined with and without this assumption. Then the different results are compared.

– The inclusion of the correlation (intrasystem) between the differents aggregated impacts of the used processes in the LCA simplified software when evaluating the impacts uncertainties of the whole system. The importance of the covariance in the calculation of variance is quantified.

– The inclusion of the correlation,(intersystem) comparing two system of the aggregated products in comparaison of the impacts of two systems of products to determine the propability that one will be better than the other.

– Realizing a proof of concept applying the research to a case study with an industrial partener. The used tool has to meet the new manufacturers' expectations with a new technology which requires less time, less energy, less space and less labour.

The methodology constitutes to evaluate the three correlation levels regarding the propagation uncertainties, and to consider the results that the tool brings.

The research evaluates the propagation impacts is not really sensitive to the hyposthisis of the intradata correlation. However, the intrasystem correlation changes considerably the results of the impacts uncertainties of an aggregated product system. In addition, the developed tool allows to meet the industrial expecations in terms of simplicity, time and clarity. It is possible to integrate the uncertainties information to simplified LCA softwares considering the different levels of the correlation see in the research.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

ICV	Inventaire du Cycle de Vie
ÉICV	Évaluation des Impacts du Cycle de Vie
BD ICV	Bases de Données des Inventaires du Cycle de Vie
ACV	Analyse de Cycle de Vie
LCA	Life Cycle Assessment
CC	Changement climatique (Climate Change)
HH	Santé humaine (Humain Health)
EQ	Qualité de l'écosystème (Ecosystem Quality)
R	Ressources (Resources)
WW	Prélèvement de l'eau (Wather Withdrawal)
E-4	10^{-4}
$\mathbf{M}^\top$	transposée de la matrice $\mathbf{M}$
$ln(x)$	logarithme néperien de x
$e^x = exp(x)$	exponentielle de x
$X \sim LN(\mu, \sigma)$	X suit la loi Log-Normale de paramètres μ et σ
$X \sim N(\mu, \sigma)$	X suit la loi Normale de paramètres μ et σ
$\mathrm{E}(X)$	Espérence mathématiques de la variable aléatoire X : $\mathrm{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$
$\mathrm{Var}(X)$	Variance de la variable aléatoire X : $\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]$
$\mathrm{Cov}(X, Y)$	Covariance entre les v.a X et Y : Cov(X,Y) = $\mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$
$C_v(X)$	Coefficient de variation de la v.a X : $C_v(X) = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)}$

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

1.1 Contexte

Toute mesure vient nécessairement avec une incertitude. Cette incertitude doit être prise en compte dans l'interprétation des mesures, en particulier dans le domaine du développement durable, celui qui sera étudié ici. Lorsque des informations portant sur des résultats d'impacts environnementaux sont communiquées, il faut être conscient que ces données présentent une incertitude et que les conclusions qui vont être tirées en dépendent directement. L'étude réalisée portera plus précisément sur l'incertitude des résultats issue de la méthode de l'ACV. Cette méthode permet d'évaluer les impacts d'un produit, d'un procédé ou d'un service en considérant l'intégralité de son cycle de vie. L'ACV se base sur la mesure des entrants et sortants des procédés qui composent un cycle de vie. Par la suite la manière dont ces différents procédés font appel les uns aux autres est représentée sous la forme d'un graphe. À l'aide de ce modèle ACV il est possible d'obtenir des résultats sur les impacts environnementaux du cycle de vie d'un produit analysé. Nous nous intéresserons ici à la propagation des incertitudes à travers le modèle de l'analyse de cycle de vie. Plus précisément, nous cherchons à connaître les incertitudes des impacts d'un cycle de vie complet à partir des incertitudes des impacts des processus isolés et les incertitudes sur la manière dont ils sont liés les uns aux autres.

Les incertitudes en ACV apparaissent à de multiples niveaux. Tout d'abord les valeurs correspondant aux émissions ou aux interactions mesurées sont stockées dans des bases de données et présentent une certaine incertitude liée à la qualité des mesures (localisation, instruments, variabilité ...). Ensuite lors de la construction du modèle représentant les interactions des processus dans le cycle de vie su système, les choix méthodologiques ou de données sont aussi sources d'incertitudes. Les hypothèses sur les corrélations entre tous les flux et les procédés intervenants dans un cycle de vie sont au cœur de l'étude.

Actuellement, pour des raisons de simplicité et de rapidité les industries utilisent des logiciels ACV simplifiés. Ces logiciels utilisent des données agrégées correspondant directement aux impacts du cycle de vie complet de systèmes élémentaires et sont dépourvus d'informations sur les incertitudes des résultats. Cette étude vise donc à pouvoir intégrer des informations sur les incertitudes des résultats dans les logiciels ACV simplifiés. Une série de contraintes est donc à prendre en compte, comme l'utilisation de données pré-calculées (dites agrégées) ne donnant pas accès au cycle de vie complet des procédés, la prise en compte de différents niveaux de corrélations entre les différents processus intervenant dans le cycle de vie du système, ou encore des exigences de clarté des résultats et de rapidité d'exécution des calculs.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 ACV

2.1.1 Contexte ACV

L'Analyse de Cycle de Vie est une méthode principalement utilisée en développement durable pour évaluer et comparer les impacts environnementaux du cycle de vie complet d'un produit ou d'un service. Toutes les étapes de la vie du produit doivent être considérées, l'extraction des matières servant à la production, la production, l'utilisation et la fin de vie ou le recyclage. Les relations entre ces différentes étapes du cycle de vie sont représentées dans la Figure 2.1.

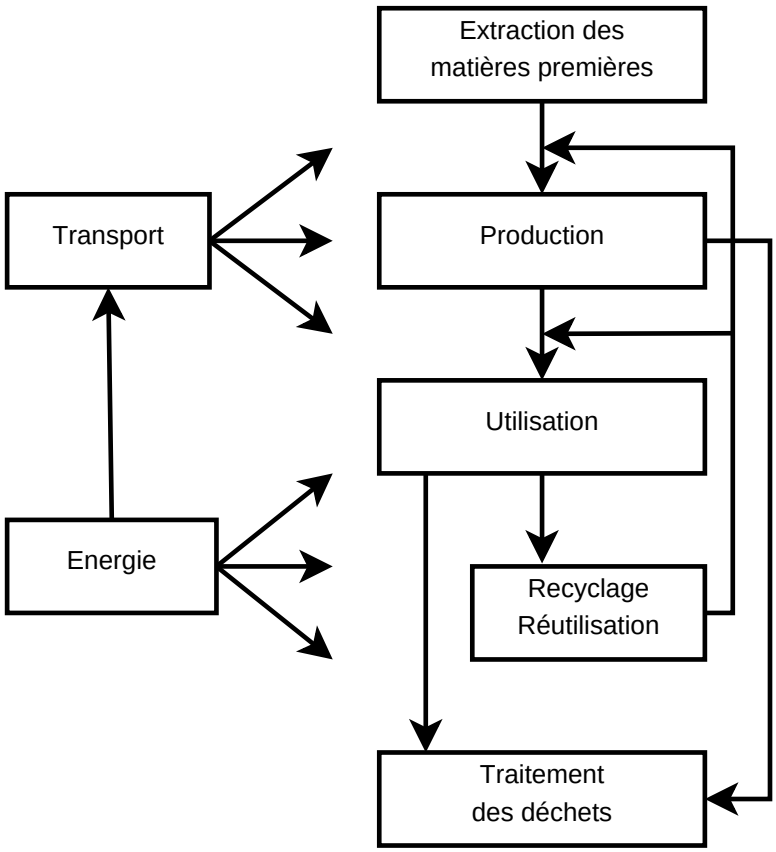


Figure 2.1 Étapes principales d'un cycle de vie.

L'ensemble des de ces étapes constitue le système de produit. Il représente les impacts du produit allant du « berceau » au « tombeau ». La méthode est documentée dans les normes ISO14040 (ISO, 2006a) et ISO14044 (ISO, 2006b) selon lesquelles la méthode ACV doit suivre les quatre étapes majeures suivantes :

1. « Définition des objectifs et du champ de l'étude » : cette étape définit les objectifs, les limites et le champ de l'étude. C'est ici que l'unité fonctionnelle du produit ou du service est définie. Cette unité fonctionnelle représente la demande exogène du système de produit, et les résultats de l'étude doivent se ramener à cette unité de demande.
2. « Inventaire du Cycle de Vie (ICV) » : Cette étape nous permet de comptabiliser l'intégralité des entrants et sortants du système de produit rapportés à l'unité fonctionnelle. Ces entrants et sortants sont composés de flux élémentaires. Ces flux correspondent aux émissions (sortants) ou prélèvements (entrants) du système de produit dans l'écosystème. Ils sont présentés à la section 2.1.2
3. « Analyse de l'impact (ÉICV) » : Cette étape évalue l'impact environnemental potentiel des flux élémentaires entrants et sortants comptabilisés à l'étape d'inventaire en utilisant des relations de cause à effet.
4. « Interprétation » : Cette étape fait l'analyse des résultats obtenus à l'étape précédente. C'est ici que les limites de l'étude sont définies et que les objectifs fixés au début de l'étude sont validés ou révisés.

Les relations entre ces différentes étapes de l'analyse du cycle de vie selon la norme ISO 14040 sont présentées dans la Figure 2.2. On voit que l'étape d'inventaire est un intermédiaire qui permet de passer de la « Définition des objectifs et du champ de l'étude » à l' « Évaluation des Impacts du Cycle de Vie ». En revanche, l'étape d'interprétation est en relation avec les trois autres étapes, car elle est transversale. En effet, l'analyste doit constamment interpréter les données, les modèles, et les résultats présents dans les autres étapes.

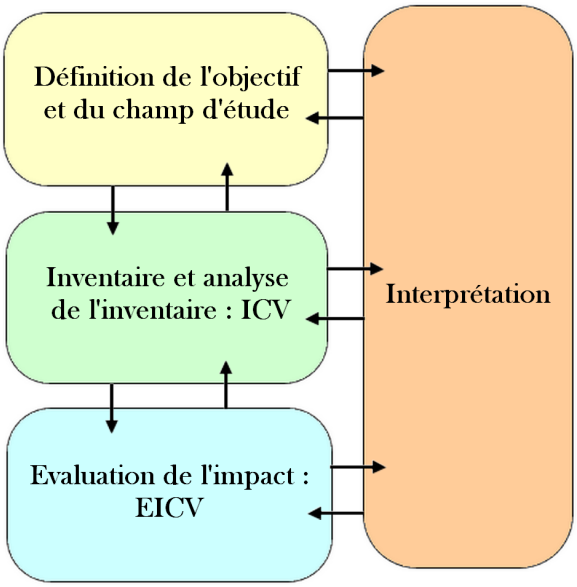


Figure 2.2 Étapes principales de l'ACV selon ISO.

Les différentes étapes sont obligatoires selon la norme ISO 14044. Cependant, il y a des sous-étapes facultatives qui peuvent s'ajouter à l'étude. Par exemple la normalisation des impacts, leur regroupement et leur pondération, ou encore l'analyse de la qualité des données et l'incertitude des résultats. C'est plus précisément à cette étape d'analyse d'incertitude qui sera étudiée par la suite. Pour cela, les étapes 2 (« Inventaire et analyse de l'inventaire») et 3 (« Analyse de l'impact (ÉICV)») seront étudiées en détaille, car elles sont nécessaires à la compréhension de l'étude.

2.1.2 Structure de calcul ACV

Pour la réalisation d'une ACV, la première étape présentée dans la section 2.1.1, consiste entre autres à définir une unité fonctionnelle, par exemple : « Produire une plaque d'acier de $1m^2 sur 0.5\text{cm} d'épaisseur$ ». Cette unité fonctionnelle représente la demande exogène, et donc le bien ou service que le système de produit doit mettre à disposition. Ensuite, pour déterminer l'inventaire du cycle de vie, il faut d'abord répertorier les produits et services nécessaires à la réalisation de cette unité fonctionnelle. Pour cela, il faut définir le processus élémentaire associé. Un processus élémentaire correspond à une activité comme de l'extraction ou de la production. Il peut être représenté par une boîte noire de laquelle certaines quantités de produits ou services vont sortir (production) et dans laquelle des produits vont entrer (consommation). Ces flux qui lient les processus élémentaires entre eux sont appelés des flux économiques. Les différents processus élémentaires seront alors liés les uns aux autres par le

biais de ces flux économiques. Les émissions et prélèvements dans la nature y sont représentés et sont appelés flux élémentaires. La Figure 2.3 présente un processus élémentaire générique.

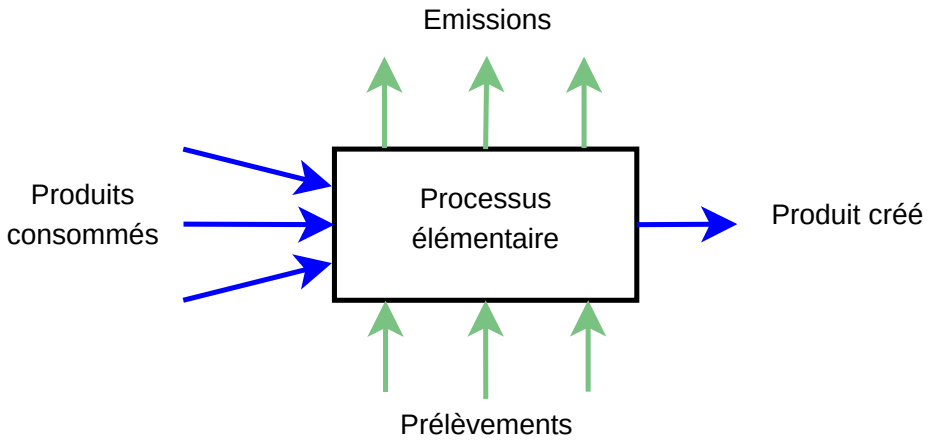


Figure 2.3 Processus élémentaire avec les flux économiques (en bleu) et les flux élémentaires (en vert) associés.

Toujours avec l'exemple précédent, le processus élémentaire de production d'une plaque en acier de $1m^2$x$0.5cm$ nécessite une certaine quantité d'acier (produit) ainsi qu'un aplatissement (service). Des processus élémentaires fournissant tous les produits et services sont modélisés et reliés afin de constituer un cycle de vie. Les services qui nécessitent des matières entrantes (ici de l'acier) pour les transformer en produits sortants (ici une plaque de $1m^2$x$0.5cm$ d'acier) nécessitent en plus certaines quantités de produits et services dont ils ont besoin pour être réalisés (énergie, autre plaque d'acier pour écraser...), et interagissent avec la nature avec leurs émissions et prélèvements dans l'environnement (Frischknecht et al., 2007). De cette manière va être constitué un arbre qui représente le cycle de vie correspondant à notre unité fonctionnelle. Ces arbres sont présentés en détail dans la section 2.1.5, car au cœur de la problématique de propagation des incertitudes.

Lors de l'inventaire, l'unité fonctionnelle est représentée par un vecteur de demande finale : $\mathbf{f}$. Ce vecteur de demande finale représente la demande exogène du système de produit. Il « appelle » les biens et services directement nécessaires pour fournir l'unité fonctionnelle (Rebitzer et al., 2004). Dans l'exemple suivant, il correspond à un appel d'une plaque d'acier de $1m^2$, associé au processus élémentaire de production de plaques d'acier.

Une fois le vecteur de demande finale défini, il reste à déterminer les quantités des produits et services (flux économiques) associés au cycle de vie complet de notre unité fonctionnelle.

Pour cela, un graphe (arbre/réseau) représentant l'intégralité du cycle de vie est construit, cela correspond au diagramme des flux de procédés présenté par Suh et Huppes (Suh and

Huppes, 2005). Chaque nœud correspond à un processus élémentaire monofonctionnel, c'est à dire avec comme seul produit d'intérêt le flux économique associé. Chaque liaison entre deux processus élémentaires correspond à un flux économique. Une rétroaction est appliquée sur les processus élémentaires qui sont déjà présents. Ensuite, la matrice de structure correspondant à ce graphe est construite. Les colonnes correspondent aux processus élémentaires monofonctionnels, et les lignes aux flux économiques associés, cette matrice est donc carrée. Chaque colonne correspond à un processus élémentaire (production d'un certain produit ou service) noté j, et chaque élément i de la colonne j correspond à la quantité du produit ou de service i qui est consommée ou créé lors de la production du processus de la colonne j. Si la quantité du produit est consommée alors la valeur est négative et si la quantité de produits est créée alors la valeur est positive. En général, les éléments diagonaux sont positifs, car correspondants à la quantité de produits que doit créer le processus élémentaire de ce même produit. Et les éléments non diagonaux sont généralement négatifs (quand ils ne sont pas nuls), car ils correspondent aux quantités des autres produits consommés par un processus élémentaire. Cette matrice est appelée la matrice technologique et est notée **A**.

Un exemple fictif et simplifié est donné avec 4 processus élémentaires :

- Production d'une plaque d'acier de $1m^2$x0.5cm :
  - consomme 10kg d'acier
  - consomme un service de laminage de 10kg d'acier en une plaque de $1m^2$x0.5cm
  - produit une plaque d'acier de $1m^2$
- Production 1kg d'acier
  - produit 1kg d'acier
- Laminage de 10kg d'acier en une plaque de $1m^2$x0.5cm en utilisant une plaque d'acier
  - consomme 10^{-6} plaques d'acier pour compresser (1 plaque sert 10^6 fois et donc produit 10^6 plaques)
  - consomme 0.5kWh d'électricité
  - produit la transformation de 10kg d'acier en une plaque de $1m^2$x0.5cm en acier
- Production 1Wwh d'électricité
  - produit 1kWh d'électricité

La structure associée aux processus et la matrice technologique associée à cette base de données seraient les suivantes :

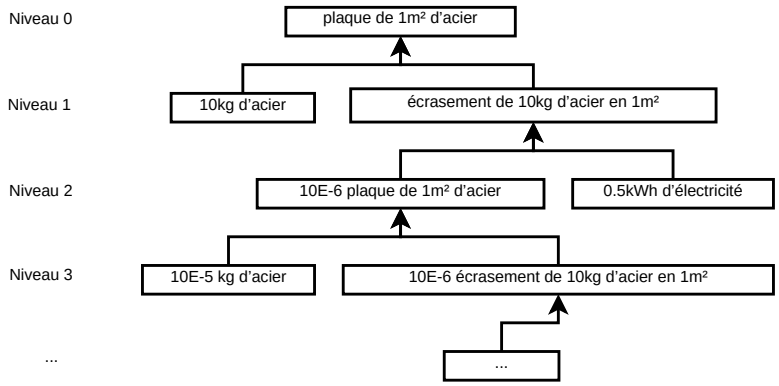


Figure 2.4 Exemple de système de produit simple et fictif pour la production d'une plaque de $1m^2$x0.5cm d'acier.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -10^{-6} & 0 \\ -10 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & -0.5 & 1 \end{pmatrix} \qquad \text{Base de } \mathbf{A} = \begin{pmatrix} m^2 \text{ de plaque d'acier} \\ \text{kg d'acier} \\ m^2 \text{ de laminage d'acier} \\ \text{kWh d'électricité} \end{pmatrix}$$

Dans la pratique, les systèmes de produits contiennent beaucoup plus de processus élémentaires, de l'ordre de quelques centaines à quelques milliers.

Dans cet exemple simple, le vecteur de demande serait $\mathbf{f} = (1, 0, 0, 0)$. Le vecteur $\mathbf{f}$ représente les quantités des produits qui sont nécessaires pour produire le « niveau 0», c'est-à-dire la demande qui permet de produire directement l'unité fonctionnelle. Cependant, ce n'est pas suffisant pour décrire le cycle de vie complet, car il faut considérer les « niveaux suivants ». Le vecteur $(\mathbf{I} - \mathbf{A})\mathbf{f}$ peut être lui-même considéré comme le vecteur de demande du niveau suivant. C'est-à-dire les quantités qui ont été utilisées pour produire les processus utilisés au « premier niveau ». Et ainsi de suite jusqu'à l'infini. Cela amène à la série entière suivante (Suh and Heijungs, 2007) :

$$\mathbf{f} + (\mathbf{I} - \mathbf{A})\mathbf{f} + (\mathbf{I} - \mathbf{A})(\mathbf{I} - \mathbf{A})\mathbf{f} + ... = \Big(\sum_{\mathbf{n=0}}^{\infty} (\mathbf{I} - \mathbf{A})^{\mathbf{n}} \Big)\mathbf{f} = \mathbf{A^{-1}f} = \mathbf{s}$$

Une étude sur la validation des conditions sur le rayon spectral de la matrice technologique assure la convergence de cette série entière et l'invisibilité de cette matrice technologique (Suh and Heijungs, 2007).

De cette manière, les productions de tous les processus élémentaires du système de produit correspondant à l'intégralité du cycle de vie de notre unité fonctionnelle sont cumulées et

rapportées à cette unité fonctionnelle. Le vecteur $\mathbf{s}$ obtenu est appelé « vecteur de mise à l'échelle ».

Il est possible de raisonner dans l'autre sens, et de se demander quel est le vecteur de mise à l'échelle qui permettrait de répondre à l'unité fonctionnelle. Autrement dit, déterminer le vecteur qu'il faut multiplier à la matrice $\mathbf{A}$ pour obtenir le vecteur de demande $\mathbf{f}$.

$$\mathbf{As} = \mathbf{f} \quad \text{d'où} \quad \mathbf{s} = \mathbf{A^{-1}f}$$

Dans cette équation, on est assuré que la matrice technologique $\mathbf{A}$ est inversible car c'est la forme déjà imputée qui est utilisée.

Chaque processus élémentaire est associé une série d'émissions et de prélèvements dans l'environnement (flux élémentaires) comme on peut le voir dans la Figure 2.3. Pour obtenir les flux élémentaires correspondants à l'intégralité du cycle de vie, il faut multiplier les quantités des processus élémentaires mis à l'échelle de l'unité fonctionnelle par leurs émissions dans l'environnement.

Cette opération peut être faite directement par un produit matriciel. Pour cela, la matrice intervention notée $\mathbf{B}$ qui correspond aux émissions et aux extractions est construite de la manière suivante. Chaque colonne du cette matrice correspond à un processus (les mêmes que dans la matrice technologique et dans le même ordre) et chaque ligne correspond à un flux élémentaire.

Le résultat précédent est alors multiplié par cette matrice $\mathbf{B}$ afin d'obtenir la totalité des émissions correspondant à l'unité fonctionnelle (Heijungs et al., 2005).

$$\mathbf{g} = \mathbf{Bs} = \mathbf{BA^{-1}f}$$

Où $\mathbf{g}$ correspond à l'inventaire du cycle de vie. C'est-à-dire l'ensemble des flux élémentaires du cycle de vie complet du système. Le vecteur $\mathbf{g}$ a autant d'éléments qu'il y a de flux élémentaires considérés, et les valeurs correspondent aux quantités de chaque flux élémentaire, normalisé par l'unité fonctionnelle, échangé entre le système de produit et l'environnement pendant le cycle de vie du produit.

Il reste alors à calculer la contribution des émissions dans différentes catégories d'impacts comme les effets cancérigènes, l'occupation des sols, etc. Pour cela, une matrice appelée matrice des facteurs caractérisation, notée $\mathbf{C_F}$, est utilisée. Chaque ligne correspond à une catégorie d'impact, et chaque colonne correspond aux contributions d'une émission dans ces catégories. Ces facteurs caractéristiques permettent de caractériser les impacts potentiels

des flux élémentaires dans les catégories d'impacts ou de dommage respectives. (Goedkoop et al., 2009; Goedkoop and Spriensma, 2001; Jolliet et al., 2003). L'équation finale permettant d'obtenir les impacts à un niveau de dommage (vecteur d'impact noté **h** dans lequel chaque élément correspond aux impacts dans une catégorie de dommage) à partir des éléments définis précédemment :

$$\mathbf{h} = \mathbf{C_F^\top g} = \mathbf{C_F^\top B A^{-1} f}$$

2.1.3 Base de données en ACV

Les bases de données en ACV contiennent de l'information sur des processus élémentaires, notamment les flux économiques et flux élémentaires qui leur sont associés. En effet, il serait impossible de peupler à chaque étude l'intégralité du système de produit correspondant au cycle de vie complet. C'est pourquoi des bases de données sont constituées à partir d'informations issues d'entreprises, par exemple et permettent de compléter les données des processus unitaires intervenant dans la matrice technologique et des flux élémentaires de la matrice d'intervention.

Ces bases de données ne contiennent pas seulement les valeurs des flux. Elles contiennent aussi certaines « métadonnées » donnant des informations sur la provenance et la qualité des données. Dans les bases de données **ecoinvent**, il y a notamment des informations permettant de déterminer l'incertitude du flux (Frischknecht et al., 2007). La base de données utilisée dans l'étude est « **ecoinvent v2.2** ». Cette base de données contient 4161 processus et 1615 types émissions.

En annexe A se trouve une description précise des valeurs de la base de données : distributions des variables, paramètres.

Le centre **ecoinvent** a publié en 2013 la version 3 de la base de données **ecoinvent**. Cette base contient plus de 10 000 processus élémentaires. De plus, il existe d'autres fournisseurs de bases de données comme **GaBi** ou **Agri-balise** qui ne seront pas présentés ici.

2.1.4 Cas des données agrégées

Les données agrégées sont des données d'inventaire du cycle de vie qu'on trouve dans les Bases de Données des Inventaires du Cycle de Vie (BD ICV). Ces données agrégées contiennent les flux élémentaires associés au cycle de vie entier d'un produit ou service (Heijungs and Suh, 2002). Les flux élémentaires d'une donnée agrégée i représentent les flux élémentaires obtenus après le calcul de l'inventaire du cycle de vie g pour un vecteur de demande correspondant à

une unité du processus i.

L'utilisation de ces données agrégées permet de réduire la taille de la base de données : la matrice technologique et la matrice intervention disparaissent et il ne reste qu'une matrice de la taille de la matrice intervention. Cette matrice est appelée la matrice intensité, est notée $\mathbf{\Lambda}$ et se calcul comme suite :

$$\mathbf{\Lambda = BA^{-1}}$$

Cependant, la matrice intensité n'est absolument pas creuse. Car pour chaque processus élémentaire (colonne) les flux élémentaires du cycle de vie complet sont exhaustifs. Alors que dans la base de données classique, la matrice technologique et la matrice intervention sont extrêmement creuses.

Un avantage réel de l'utilisation de données agrégées est l'économie du calcul. En effet, en passant par un système de produit composé de données agrégées, le calcul de l'inventaire revient à sommer tous les flux élémentaires (agrégés) des processus, puis les mettre à l'échelle de l'unité fonctionnelle. C'est pourquoi les logiciels ACV simplifiés utilisés dans l'industrie se basent généralement sur ce type de données qui permettent plus de simplicité et de rapidité de calcul (Rice et al., 1997).

2.1.5 Graphes vs réseaux en ACV

L'exemple de la figure 2.4, montre que le cycle de vie d'un produit peut être représenté sous la forme d'un arbre. La racine de l'arbre contient le processus élémentaire fournissant l'unité fonctionnelle. Viendront ensuite au 1er niveau les processus élémentaires qui contribuent directement à la production des processus élémentaires du niveau 0, et ainsi de suite. Chaque nœud de l'arbre représente un processus élémentaire, et ses « enfants » représentent les processus élémentaires qui produisent les flux économiques dont il a besoin pour être produit, ce sont les éléments de la matrice technologique ($\mathbf{a_{ij}}$).

Il est important de noter que l'arbre sera généralement infini. Cela vient du fait que dans la réalité, chaque activité (processus élémentaire) a nécessairement au moins un flux entrant. Par exemple le transport : un processus de transport aura besoin d'un processus véhicule (un camion), et ce véhicule aura besoin de matériaux qui eux-mêmes devront être transportés par un processus de transport (camion). Ainsi, une boucle infinie est créée(Arthur et al., 1987).

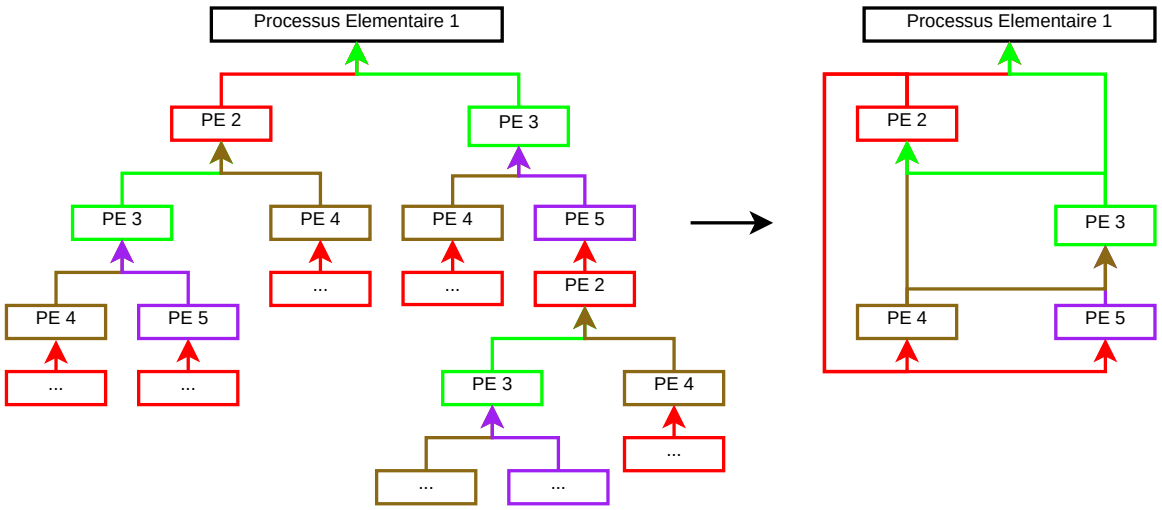


Figure 2.5 Exemple de graphe et son réseau associé.

C'est pourquoi certains logiciels vont transformer l'arbre en réseau avec des rétroactions des processus qui apparaissent déjà dans l'arbre comme le présente figure 2.5. De cette manière, le graphe est nécessairement fini avec un nombre de nœuds maximums égal au nombre de processus dans la matrice **A**. C'est de cette manière que la matrice technologique est construite. Cette dernière correspond à la matrice de structure reliant tous les processus de la base de données.

En termes de méthodes de calcul de l'inventaire de cycle de vie, les représentations d'arbre et de réseau correspondent respectivement à la série entière et la matrice inverse. En effet, chaque terme de la série entière correspond aux contributions des flux économiques d'un niveau de l'arbre, ce qui est en accord avec l'aspect infini de l'arbre pour converger vers la réalité. D'un autre côté, dans la matrice inverse les éléments d'une colonne correspond aux contributions du cycle de vie complet (mis à l'échelle) du processus élémentaire associé à la colonne, tout comme le réseau où toutes les contributions d'un produit au cycle de vie sont rebouclées et sommées.

2.2 Incertitudes en ACV

2.2.1 Sources des incertitudes en ACV

En se basant sur des études qui énoncent les différents types et les différentes sources d'incertitudes intervenant dans l'analyse de cycle de vie, il est possible d'identifier trois types principaux : les incertitudes paramétriques, les incertitudes de modélisation et les autres (Ciroth et al., 2004; Heijungs and Lenzen, 2014; Huijbregts, 1998; Lloyd and Ries, 2007).

Incertitudes paramétriques

L'incertitude paramétrique se trouve au niveau des valeurs isues de la réalité. Cette incertitude correspond à la différence entre la réalité et la valeur utilisée dans le modèle. Elle peut être due à la méthode de mesure utilisée, à l'instrument de mesure utilisé, à l'imprécision de la personne qui applique la mesure, à une quantité insuffisante des données, à une mauvaise correspondance géographique ou temporelle de la donnée utilisée ou à la variabilité intrinsèque de la donnée. Il est important de remarquer que c'est le type d'incertitude qui est le plus facile à quantifier.

Incertitudes de modélisation

L'incertitude de modélisation découle du fait que par définition un modèle est une imitation imparfaite de la réalité. Cette incertitude est intrinsèque au modèle et peut provenir de plusieurs sources comme les approximations mathématiques du modèle, les hypothèses simplificatrices du modèle, exclusion ou substitution de variables.

Autres incertitudes

On trouve principalement deux autres types d'incertitudes, l'incertitude de décision, qui vient du choix du modèle ou des variables qui ne peut pas être fait parfaitement. Et enfin, on trouve aussi les incertitudes épistémiques, qui sont causées par le manque de connaissance de la réalité. Ces deux types d'incertitudes peuvent être réintégrés dans les incertitudes paramétriques en considérant le choix restreint des variables d'entrées que l'on connaît et dans les incertitudes de modélisations par le choix du modèle qui ne peut représenter parfaitement la réalité.

Dans le cadre de l'étude réalisée, ce sont principalement les incertitudes de paramètre qui vont être traitées. En effet, la propagation des incertitudes se base sur les incertitudes des flux qui sont des incertitudes paramétriques. Ensuite, les incertitudes de modélisation interviennent dans le calcul de la propagation des incertitudes au travers des hypothèses ou des approximations des modèles analytiques ou statistiques utilisés. Par exemple, l'étude réalisée analyse entre autres la sensibilité d'une hypothèse lors du calcul de la propagation de l'incertitude lors des simulations de Monte-Carlo, il est question ici d'une incertitude de modélisation.

2.2.2 Incertitudes dans les bases de données ecoinvent

Comme il a été mentionné dans les parties précédentes, les informations sur les incertitudes sont renseignées au niveau des flux contenus dans la base de données utilisée. Ces informations sont présentées comme le « pedigree » des données (Muller et al., 2014). Plus particulièrement, les informations sur l'incertitude paramétrique au niveau des flux sont collectées dans la base de données ecoinvent et sont présentées sous la forme suivante :

- La loi de distribution du flux : Log-normale, Normale, Triangulaire...
- La valeur déterministe : elle est directement liée à un des paramètres de la distribution.
- L'écart type de la distribution : Il sera directement lié à l'autre paramètre de la loi (dans le cas de la loi triangulaire, on trouvera 2 paramètres qui seront le minimum et le maximum)
- Le pedigree : Il correspond à l'origine des incertitudes et permet de déterminer l'écart type de la variable aléatoire du flux. Il contient de plus la manière dont le flux a été déterminé (estimation, expérience ...)

Ce sont les valeurs de la matrice pedigree qui permettent de déterminer l'écart type de la distribution correspondant à la variable aléatoire du flux.

Les valeurs de la matrice pedigree sont les suivantes :

- U_1 : Facteur de fiabilité
- U_2 : Facteur d'exhaustivité
- U_3 : Facteur de corrélation temporelle
- U_4 : Facteur de corrélation spatiale
- U_5 : Facteur de corrélation technologique
- U_6 : Facteur lié à la taille de l'échantillon

Un facteur de base supplémentaire U_b est considéré selon le type de donnée. Ce facteur d'incertitude de base correspond à une variable aléatoire de départ pour le flux qui est ensuite modifiée par les informations du pedigree. Cette variable est spécifique au type de donnée collectée. Par exemple, en général les émissions de CO_2 ont des incertitudes supposément moins grandes que celles de CO (Frischknecht et al., 2007).

Chaque paramètre dans la base de données ecoinvent se voit attribuer une note comprise entre 1 et 5 pour les 6 facteurs. S'il n'y a pas assez d'informations pour délivrer une note au facteur, la note maximale de 5 se verra attribuée. Cette note va correspondre à une valeur du facteur assimilable à un écart type spécifique selon le Tableau 2.1, le Tableau 2.2 précise les critères de cette notation.

La formule permettant de déterminer l'écart type géométrique de la variable aléatoire à partir des facteurs de la matrice pedigree est la suivante :

$$GSD = e^{\sqrt{(\ln(U_1))^2+(\ln(U_2))^2+(\ln(U_3))^2+(\ln(U_4))^2+(\ln(U_5))^2+(\ln(U_6))^2+(\ln(U_b))^2}}$$

L'écart type géométrique est utilisé ici car il permet de déterminer directement l'intervalle de confiance à 95% dans le cas d'une loi log-normale, ce qui est souvent le cas ici. La loi de la distribution est renseignée dans la base de données et les paramètres de cette loi sont déterminés à partir de la valeur déterministe et de cet'écart type géométrique.

Dans la majorité des cas, la distribution des flux est log-normale. En effet, les propriétés de la loi de probabilité Log-normale se retrouvent dans les caractéristiques des flux, comme la probabilité que la variable soit négative est nulle, ce qui correspond au fait qu'un flux qui est censé être positif, ne sera jamais négatif (Limpert et al., 2001). Une loi Log-normale est caractérisée par ses deux paramètres μ et σ qui représentent respectivement la moyenne et l'écart type de la loi normale associée qui est définit comme suite : si X suit une loi normale de moyenne μ et d'écart type σ, alors $Y = e^X$ suit une loi log-normale de paramètres μ et σ. Les paramètres de la loi log-normale peuvent être estimés par la méthode des moments appliquée à une série de variables observées $\tilde{Y}$ de la variable aléatoire Y grâce aux formules suivantes :

$$Y \sim LN(\hat{\mu}, \hat{\sigma}) \text{ avec } \begin{cases} \hat{\mu} = \ln(\mathrm{E}(\tilde{Y})) - \frac{1}{2}\ln(1 + \frac{\mathrm{Var}(\tilde{Y})}{(\mathrm{E}(\tilde{Y}))^2}) \\ \hat{\sigma^2} = \ln(1 + \frac{\mathrm{Var}(\tilde{Y})}{(\mathrm{E}(\tilde{Y}))^2}) \end{cases}$$

Tableau 2.1 Valeurs des facteurs de la matrice pedigree selon leur note.

Note	1	2	3	4	5
Fiabilité	1.00	1.05	1.10	1.20	1.50
Exhaustivité	1.00	1.02	1.05	1.10	1.20
Corrélation temporelle	1.00	1.03	1.10	1.20	1.50
Corrélation géographique	1.00	1.01	1.02	-	1.10
Corrélation technologique	1.00	-	1.20	1.50	2.00
Taille de l'échantillon	1.00	1.02	1.05	1.10	1.20

Tableau 2.2 Matrice pedigree utilisée pour déf nir la qualité des données, adapté de (Frischknecht et al., 2007).

Indicateur	1	2	3	4	5	Remarques
Fiabilité	Donnée mesurée et vérif ée	Donnée d'hypothèse et vérif ée ou donnée mesurée et non vérif ée	Donnée estimée et non vérif ée	Estimation ou donnée théorique	Estimation non qualif ée	Vérif ée : publiée of ciellement
Exhaustivité	Donnée représentative : 100% des sites	Donnée représentative : > 50% des sites	Donnée représentative : < 50% des sites ou > 50% des sites sur une période plus courte	Donnée représentative : 1 seul site	Représentativité inconnue	Site : site pertinent pour le marché considéré sur une période adéquate pour uniformiser les f uctuations normales
Corrélation temporelle	< 3 ans après 2000	< 6 ans après 2000	< 10 ans après 2000	< 15 ans après 2000	> 10 ans après 2000 Ou âge inconnu	L'année 2000 est l'année de référence.
Corrélation géographique	Donnée de la zone d'étude	Donnée moyennes sur une zone plus large (intégrant la zone d'étude)	Donnée d'une zone plus petite que la zone d'étude ou d'une zone similaire	-	Donnée d'une zone inconnue ou signif cativement dif érente	La similarité est exprimée en terme législatif
Corrélation technologique	Donnée du procédé/matériau étudié et même technologie	-	Donnée de procédé/matériau approchant et de même technologie ou donnée de procédé/matériau de l'étude et de technologie dif érente	Donnée de procédé/matériau approchant et de technologie dif érente ou donnée obtenue en laboratoire et de même technologie	Donnée de procédé/matériau approchant et obtenue en laboratoire et de technologie dif érente	Exemple de technologies Différentes : turbine à vapeur au lieu d'un moteur à propulsion dans un navire
Taille de l'échantillon	> 100, mesure continue, bilan des produits achetés	> 20	> 10, donnée agrégée dans un rapport environnemental	>= 3	Inconnue	La taille de l'échantillon d'une donnée dans sa source d'information

2.2.3 Propagation des incertitudes en ACV

La propagation de l'incertitude se fait lors de l'étape d'inventaire et analyse de l'inventaire. En effet, la majorité des éléments de la matrice technologique et de la matrice inventaire sont des variables aléatoires. Le résultat des impacts est fonction de ces deux matrices et est donc lui aussi une variable aléatoire.

$$\mathbf{h} = \mathbf{C_F^\top B A^{-1} f}$$

- $\mathbf{h}$: vecteur des impacts du cycle de vie complet de l'unité fonctionnelle de l'étude (aléatoire)

- $\mathbf{C_F}$: Matrice des facteurs de caractérisations permettant le regroupement des émissions en catégories (aléatoire, cependant aucune information sur les incertitudes n'est pris en compte dans l'étude, donc la matrice est considérée déterministe)

- $\mathbf{B}$: Matrice (intervention) des émissions directes des processus de la base de données (aléatoire)

- $\mathbf{A}$: Matrice (technologique) des consommations directes entre tous les processus de la base de données (aléatoire)

- $\mathbf{f}$: vecteur (demande) des processus directement nécessaires à la réalisation de l'unité fonctionnelle (normalement déterministe).

L'impact sera donc fonction des éléments des matrices $\mathbf{A}$ et $\mathbf{B}$ qui sont des valeurs aléatoires, ce sera alors aussi une variable aléatoire. Cependant, on ne peut pas connaître directement l'allure de sa distribution de probabilité ou même son incertitude à partir de celles des éléments de $\mathbf{A}$ et $\mathbf{B}$. Cela vient du fait que l'opération n'est pas linéaire, effectivement nous avons affaire ici à une inversion de matrice de variables aléatoires suivit d'un produit matriciel.

Il existe cependant plusieurs méthodes permettant d'avoir une idée de l'incertitude des résultats au niveau des impacts. Celles qui sont exposées ici sont les deux classiquement utilisées.

1re méthode : méthode analytique

L'approche analytique revient à linéariser l'opération précédente en utilisant les séries de Taylor (Imbeault-Tétreault et al., 2013; Hong et al., 2010). Ensuite, avec une approximation du calcul de la variance d'une fonction de variable aléatoire, il est possible d'obtenir la variance

de l'impact analytiquement. Plus précisément en considérant deux variables aléatoires x et y, puis une autre variable aléatoire z fonction de x et de y par une fonction quelconque f, la variance de z peut être approximé par les variances de x et de y de la manière suivante (Heijungs, 2010) :

$$z = f(x, y)$$

$$\mathrm{Var}(z) = \mathrm{Var}\Big(f(x,y)\Big) \simeq \Big(\frac{\partial z}{\partial x}\Big)^2 \mathrm{Var}(x) + \Big(\frac{\partial z}{\partial y}\Big)^2 \mathrm{Var}(y) + \Big(\frac{\partial z}{\partial x}\Big)\Big(\frac{\partial z}{\partial y}\Big)\mathrm{Cov}(x)$$

Dans le cas du calcul en ACV avec pour hypothèse que les variables aléatoires présentes dans les matrices A et B sont toutes indépendantes et suivent des lois log-normales, nous avons les équations suivantes (MacLeod et al., 2002) :

$$A = (a)_{ij} \quad et \quad B = (b)_{ij}$$

$$\Big(\ln\left(GSD_h^2\right)\Big)^2 = \sum_x S_{x,h}^2 (\ln\left(GSD_x\right)^2)^2 = \sum_{i,j} S_{a_{ij},h}^2 (\ln\left(GSD_{a_{ij}}\right)^2)^2 + \sum_{k,l} S_{b_{kl},h}^2 (\ln\left(GSD_{b_{kl}}\right)^2)^2$$

$$S_{x,h} = \frac{\partial \ln(h)}{\partial \ln(x)} = \Big(\frac{\partial h}{\partial x}\Big)\Big(\frac{x}{h}\Big)$$

Ici, ce n'est pas directement la variance qui est observée, mais le GSD^2, qui est l'écart type géométrique au carré définit comme l'exponentielle de l'écart type σ de la distribution Normale sous-tendant la Log-normale : $GSD = e^{\sigma}$. Dans ces équations, il faut connaître la variance des variables aléatoires d'entrées, les a_{ij} et b_{ij} et les dérivées partielles de h en fonction de ces variables. Les dérivées sont obtenues grâce à la linéarisation de la fonction par la série de Taylor à l'ordre 1 (Osler, 1971).

2e méthode : statistique (Simulations de Monte-Carlo)

Une autre méthode pour obtenir des informations sur les incertitudes des résultats d'impacts est de faire des simulations de Monte-Carlo. Cette méthode présente l'avantage de ne pas faire d'approximation, d'informer non seulement sur la variance, mais aussi sur la distribution du résultat de l'impact. Finalement, les tirages obtenus peuvent être utilisés pour calculer les covariances et donc les corrélations entre les différents processus élémentaires.

La méthode de Monte-Carlo (Lo et al., 2005) utilisée en ACV consiste à tirer à chaque itération les valeurs de la matrice technologique et de la matrice intervention (les a_{ij} et b_{ij}) dans leurs distributions respectives et à les utiliser pour calculer le résultat d'intérêt (**g** ou **h**). Les résultats obtenus après un certain nombre de tirages donnent un échantillon du résultat

des impacts. La matrice technologique étant relativement grande, son inversion prend un temps non négligeable à chaque itération. De plus, la méthode de Monte-Carlo nécessite d'avoir un nombre suffisant d'itérations pour que les résultats obtenus soient statistiquement significatifs. La durée des itérations et le nombre d'itérations nécessaire impliquent donc une durée non négligeable d'exécution.

2.3 Corrélation

Les données utilisées en ACV sont incertaines comme il a été exposé précédemment. De plus, des données ne sont pas nécessairement indépendantes. C'est pourquoi il a été jugé intéressant de considérer les corrélations qui interviennent dans la propagation des incertitudes dans le calcul des impacts.

2.3.1 Estimation de la covariance et de la corrélation

La covariance est une forme bilinéaire symétrique qui s'applique à des couples de variables aléatoires. La covariance entre une variable aléatoire X et une autre variable aléatoire Y est la valeur de l'espérance mathématique du produit des différences entre les variables et leur propre espérance mathématique. Cette valeur représente la correspondance entre les variations des deux variables. Si les variables évoluent de manières similaires, leur covariance est positive, si elles évoluent de manière opposée, leur covariance est négative et si elles évoluent de manière indépendante, leur covariance sera nulle.

Cependant il faut être prudent aux conclusions tirées, car la réciproque est fausse, c'est-à-dire que si la covariance entre deux variables aléatoires est nulle, cela ne signifie pas nécessairement que les variables sont indépendantes (Ross, 2003).

La covariance entre ces deux échantillons de variables aléatoires X et Y indépendants de n réalisations peut être estimée sans biais par l'équation suivante :

$$\widehat{\mathrm{Cov}(X,Y)} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)\left(Y_i - \frac{1}{n}\sum_{k=1}^{n}Y_k\right)\right]$$

De manière plus générale, la matrice de variance-covariance d'un vecteur aléatoire pourra être estimée. Cette matrice est composée des covariances de tous les couples de variables du vecteur. L'élément de la ligne i et de la colonne j sera donc la covariance entre le i^e et le j^e élément du vecteur aléatoire. Les éléments diagonaux correspondent donc aux variances des éléments du vecteur aléatoire et la symétrie de la matrice est assurée par la bilinéarité.

Il est possible de normaliser la covariance. Il est alors question de corrélation notée généralement ρ. La corrélation est le rapport entre la covariance de deux variables aléatoires et le produit de leurs écarts types.

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

La corrélation est donc une mesure normalisée de la correspondance des variations de deux variables aléatoires. Elle est comprise entre -1 et 1 et est nulle lorsque la covariance est nulle.

Ainsi, les courbes de la Figure 2.6 montrent bien l'évolution des variables aléatoires les unes par rapport aux autres pour chacunes des valeurs particulières de la corrélation.

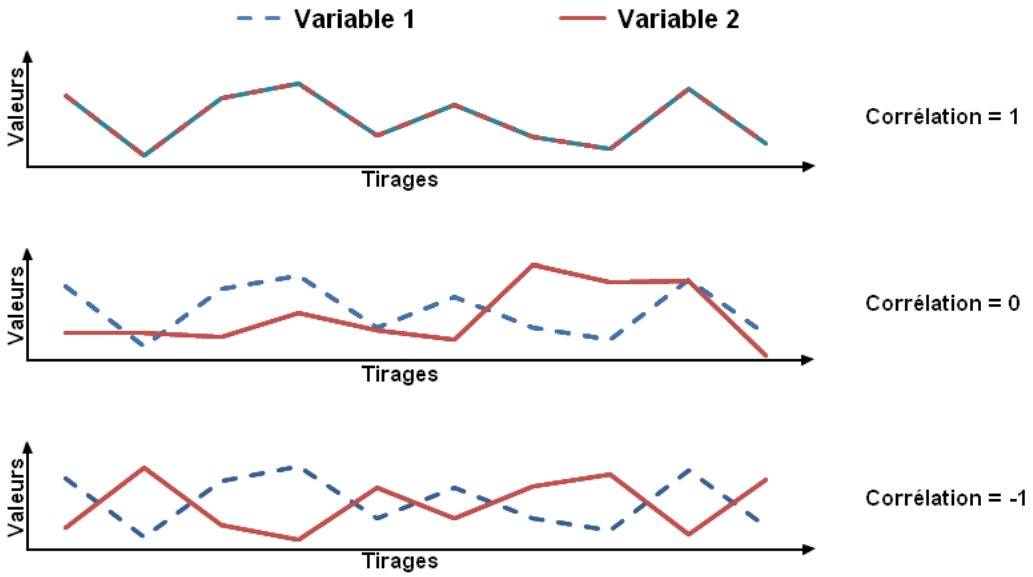


Figure 2.6 Évolution des variables aléatoires selon certaines valeurs de corrélations.

La Figure 2.6 montre que si les variations sont superposables (courbe bleue continue et rouge continue), la corrélation vaut 1, les variables sont dites parfaitement corrélées (positivement). Dans le cas où les variations sont opposées (courbe bleue continue et verte continue), les variables sont dites parfaitement corrélées négativement. Enfin, si les variations sont indépendantes (courbe bleue continue et bleue pointillée), les variables ne sont pas corrélées et la corrélation vaut 0.

Une précaution est à considérer avec la corrélation, car deux variables fortement corrélées ne signifient pas que l'une est nécessairement la cause de l'autre (Xu and Srihari). Par exemple,

deux variables aléatoires peuvent être causées par une troisième, elles seront alors corrélées entre elles sans que l'une ait causé l'autre.

2.3.2 Corrélation en ACV

Les corrélations lors de la propagation de l'incertitude en analyse de cycle de vie interviennent dans le cadre de l'incertitude de modélisation. Il existe quelques études qui considèrent les corrélations (Wei et al., 2015), mais généralement celles-ci ne sont pas considérées.

En analyse de cycle de vie, les corrélations apparaissent à plusieurs niveaux. Par exemple, il apparaît dans la corrélation entre les différents flux intervenants dans un système de produit, ou encore dans la corrélation entre deux systèmes de produits différents utilisant des flux communs (Basson and Petrie, 2007; Bojacá and Schrevens, 2010; Lo et al., 2005). En tenant compte de la contrainte de la plupart des industriels qui est l'utilisation de données agrégées dans un outil simplifié, trois niveaux principaux de corrélation schématisés dans la Figure 2.7 sont étudiés.

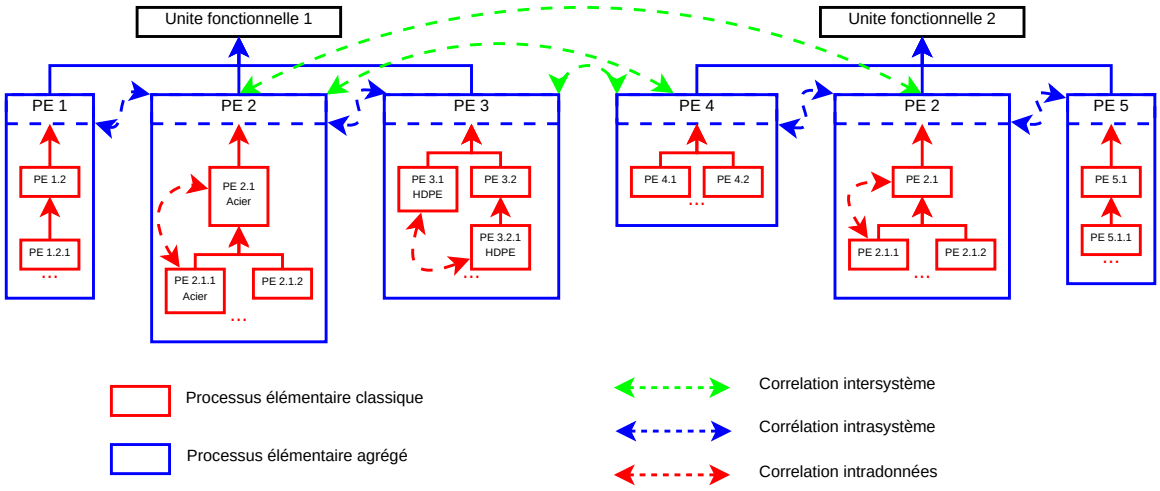


Figure 2.7 Représentation des différents niveaux de corrélation en ACV.

Corrélation intradonnées

La corrélation appelée ici « **corrélation intradonnées** » correspond à la corrélation entre un mêmes processus apparaissant à différents endroits du cycle de vie d'un produit (c'est-à-dire à différents endroits dans le système de produit qui sous-entend une donnée agrégée). Chaque impact d'un système de produit a une certaine incertitude qui est calculée avec la méthode de Monte-Carlo. Lors du calcul de cette incertitude, il y a une hypothèse de

corrélation intradonnées qui est normalement posée. Plus précisément, cette hypothèse de corrélation est faite lors du passage de l'arbre au réseau. Comme montre la Figure 2.5, le passage de l'arbre infini au réseau applique des rétroactions sur les processus déjà présents dans l'arbre (une hypothèse que chaque instance d'un processus élémentaire donné est en fait le même processus élémentaire). Ainsi la quantité nécessaire d'un même processus qui apparaît plusieurs fois dans l'arbre correspond à la somme des quantités de chaque apparition. Sommer les quantités puis tirer une seule valeur de chaque flux de la matrice **A** au lieu de tirer une valeur pour chaque apparition du flux dans l'arbre revient à considérer toutes les apparitions du flux comme totalement corrélées. C'est cette corrélation qui est notée corrélation intradonnées. Cette corrélation intervient lors du calcul des impacts agrégés des différents processus qui composent un système de produit construit exclusivement de données agrégées.

Corrélations intrasystème

La « **corrélation intrasystème** » correspond à la corrélation entre les différentes données agrégées qui composent un système de produit. Comme il a été exposé précédemment, les impacts des systèmes de produits sont corrélés, car dépendent tous des mêmes flux aléatoires, ceux présents dans la matrice technologique et la matrice intervention. Une donnée agrégée n'est autre qu'un système de produit avec un vecteur de demande unitaire (avec un 1 dans la ligne du processus que l'on souhaite agréger). Par exemple, si pour répondre à une unité fonctionnelle quelconque, un système de produits est construit avec plusieurs données agrégées (un processus d'électricité, un processus de transport, un processus d'emballage et un processus de recyclage), les impacts de ces différents processus sont corrélés, car durant leur cycle de vie, certains flux sont les mêmes. Cette corrélation est alors considérée lors du calcul des impacts du système complet.

Corrélation intersystème

La **corrélation intersystème** est du même type que la corrélation intrasystème. Elle correspond à la corrélation entre les différentes données agrégées de deux systèmes de produits comparés. Effectivement, les deux systèmes comparés sont corrélés. Il n'est alors pas correct de les considérer comme indépendants lors de leur comparaison.

2.4 Bilan de la revue de la littérature

L'analyse de cycle de vie est une méthode qui existe depuis quelques années et qui est de plus en plus utilisée dans l'industrie pour l'évaluation d'impact. La prise en compte des incertitudes dans cette méthode primordiale.

Les incertitudes au niveau des résultats d'impact sont généralement obtenues par des analyses de Monte-Carlo à partir de logiciels comme **SimaPro** ou **OpenLCA**. Au niveau de la recherche, des méthodes analytiques ont aussi été explorées (Heijungs, 2010; Heijungs and Lenzen, 2014; Hong et al., 2010). Cependant, aucune de ces deux approches ne s'est encore intéressée à analyser le niveau de corrélation intradonnées.

Dans le monde industriel, le personnel n'est pas systématiquement qualifié ou simplement n'a pas assez de temps pour réaliser des analyses de cycle de vie approfondi. Cela implique généralement l'utilisation de données agrégées qui apportent une simplicité et une rapidité des calculs des impacts dans la méthode ACV. Actuellement, l'utilisation de données agrégées ne permet pas la prise en compte des incertitudes dans l'évaluation des impacts dans les logiciels ACV simplifiés.

Finalement, il relève que les impacts de différents systèmes de produits sont corrélés. Donc dans le cas de l'utilisation de systèmes de produits agrégés ou encore dans le cas de comparaison de plusieurs systèmes de produits, les différents produits agrégés ne devraient pas systématiquement être considérés comme indépendants. Cependant, ces corrélations ne sont actuellement pas prises en compte dans les évaluations des incertitudes lors des analyses de cycle de vie.

CHAPITRE 3 PROBLÉMATIQUES ET OBJECTIFS

3.1 Problématiques

Les informations, données et modèles utilisés dans les analyses de cycles de vie sont empreints d'incertitudes, parfois considérables. Une mauvaise évaluation de ces incertitudes peut être la source d'une interprétation faussée des résultats et de la prise de mauvaises décisions. Dans les bases de données « **ecoinvent** », les informations sur l'incertitude sont disponibles seulement au niveau des flux des processus unitaires (mais pas au niveau des processus agrégés généralement utilisés dans les logiciels ACV simplifiés). Ces incertitudes peuvent être propagées afin de déterminer l'incertitude de l'inventaire grâce à des méthodes statistiques comme la méthode de Monte-Carlo. Cependant, les contraintes de temps ne permettent pas l'utilisation de telles méthodes au sein des logiciels d'ACV simplifiés.

Nous chercherons alors à savoir comment intégrer des informations d'incertitudes dans les logiciels ACV simplifiés utilisés par des industriels. Pour ce faire, il sera nécessaire de considérer au mieux les différents niveaux de corrélations pour déterminer les incertitudes des résultats finaux. De plus, l'utilisation de données agrégées dans les logiciels ACV simplifiés ne permet pas la prise en compte des corrélations entre les processus. Nous regarderons alors les différents niveaux dans lesquels les corrélations interviennent : la corrélation intradonnées entre les processus élémentaires identiques apparaissant à plusieurs reprises dans un système de produit , la corrélation intrasystème entre les processus agrégés qui composent un système de produit et la corrélation intersystème entre les processus agrégés qui composent deux systèmes de produits lors de leur comparaison. De cette problématique nous explicitons les questions de recherche suivantes :

– Une hypothèse de corrélation intradonnées est généralement faite lors du calcul de la propagation de l'incertitude. Cette hypothèse est-elle sensiblement différente à l'hypothèse dans laquelle la décorrélation est totale ? Dans le cas où la sensibilité est importante, y a-t-il des explications généralisables ? On s'attend à ce que l'hypothèse de corrélation intradonées soit sensible pour certains types de processus et donc qu'il y ait une diminution de l'incertitude dans le cas non corrélé comparativement au cas corrélé.

– Étant donné que les logiciels ACV simplifiés utilisent des données agrégées, comment évaluer l'importance de la corrélation intrasystème entre les différents processus agrégés d'un système de produit lors de l'évaluation des incertitudes associées aux impacts de ce sys-

tème ? Étant donné que les processus agrégés font très certainement appel aux mêmes flux dans leurs arrières plans, leur corrélation n'a aucune raison d'être nulle.

– Les logiciels ACV simplifiés permettent la comparaison de deux systèmes de produits. Comment considérer les corrélations intersystème entre les processus agrégés des deux systèmes de produits comparés ? Cela permettra d'intégrer une aide à la décision prenant en compte les incertitudes avec corrélations dans les logiciels. L'utilisateur pourra alors faire un choix plus éclairé.

– Les logiciels ACV simplifiés nous permettraient-ils d'intégrer des informations d'incertitudes en tenant compte de ces corrélations ? Il est ici important de tenir compte des contraintes de temps auxquelles ces outils simplifiés sont soumis et à l'importance d'une information claire et précise. De plus, serait-il souhaitable que les utilisateurs du logiciel soient capables de fixer des niveaux de corrélation préalablement connus entre les processus des différents systèmes de produits ?

3.2 Objectifs

L'objectif général de ce travail de recherche est d'intégrer l'analyse des incertitudes dans les outils ACV simplifiés qui utilisent des données agrégées en tenant compte des niveaux de corrélation entre les processus, du temps d'exécution rapide et une représentation graphique claire et précise. Il est décliné dans les objectifs spécifiques suivants :

– Développer une méthode pour quantifier la sensibilité de l'hypothèse de corrélation intra-données lors de la propagation de l'incertitude et décider de sa prise en compte. Identifier ensuite s'il y a des causes communes à la forte sensibilité de l'hypothèse puis savoir si elles sont généralisables.

– Identifier l'analyse statistique appropriée pour quantifier la corrélation (intrasystème) entre les différents processus agrégés de la base de données ecoinvent et évaluer son importance.

– Identifier une méthode pour considérer la corrélation intersystème entre les processus de différents systèmes lors de leur comparaison. Puis donner une information aidant à la décision du système le moins impactant tout en tenant compte des corrélations des différents produits.

– Coder un algorithme permettant d'intégrer dans les logiciels d'ACV simplifiés un calcul rapide de l'incertitude d'un système de produit en utilisant une base de données de processus unitaires agrégés avec leur incertitude et leurs covariances mutuelles. Puis illustrer son fonctionnement à travers une étude de cas. L'algorithme devra prendre en compte les différents niveaux de corrélation calculés précédemment et donner la possibilité à l'utilisateur de déterminer certaines corrélations connues ou d'en observer d'autres.

Le succès de ce travail de recherche portera sur la réalisation d'une preuve de concept. Ce sera un outil pour réaliser des ACV simplifiées qui devra prendre en compte les différents niveaux de corrélation considérés dans le travail de recherche. L'outil devra aussi répondre aux attentes du partenaire industriel et finalement être intégré à son logiciel d'écodesign actuellement utilisé.

CHAPITRE 4 MÉTHODOLOGIE

Le projet consiste en l'analyse de la propagation de l'incertitude lors du calcul des impacts en analyse de cycle de vie. Le point central de l'étude est de considérer les différents niveaux de corrélations exposés dans la section 3.1 lors de la propagation de l'incertitude afin de permettre les analyses d'incertitudes dans les logiciels utilisant des données agrégées. Dans une première partie, la sensibilité de l'hypothèse de corrélation « **intradonnées** » est analysée. Ensuite, la corrélation « **intrasystème** » est étudiée en comparant la part des covariances dans la variance d'un système de produits construit avec des processus agrégés. Puis la corrélation « **intersystème** » présente lors de la comparaison de deux systèmes de produits composés de processus agrégés. Finalement, un outil intégrant les analyses de l'étude est développé puis une étude de cas proposé par le partenaire Nestlé est réalisée.

4.1 Sensibilité de l'hypothèse de corrélation intradonnées

L'hypothèse de corrélation « **intradonnées** » est présente lorsqu'on passe d'un arbre infini qui correspond à la réalité de l'enchaînement des processus dans le cycle de vie d'un produit, à un réseau dans lequel la série entière est remplacée par la matrice inverse lors du calcul des impacts. Le fait d'appliquer des rétroactions sur les processus qui sont déjà présents dans le réseau en sommant les quantités à chaque retour, revient à considérer toutes ces apparitions comme totalement corrélées, car dans la simulation de Monte-Carlo, il y aura un tirage unique des variables aléatoires associées à ce processus pour la totalité des apparitions cumulées.

Dans la réalité, certaines de ces apparitions sont corrélées et d'autres non. Par exemple, un même processus de transport pour déplacer des matériaux peut intervenir plusieurs fois dans un système de produit. Si les matériaux ont été déplacés dans le même camion, alors les émissions associées au processus de transport sont totalement corrélées. Cependant, si les matériaux ont été déplacés par deux camions différents, conduits sur deux trajets différents par deux conducteurs différents, alors les émissions associées aux deux apparitions de ce même processus ne sont pas corrélées.

Lors du calcul de la propagation des incertitudes, l'hypothèse de corrélation « **intradonnées** » est normalement faite dans les logiciels ACV. De manière générale, dans les bases de données agrégées, les impacts des systèmes de produits élémentaires sont normalement calculés en faisant l'hypothèse de corrélation « **intradonnées** ».

Afin d'analyser la sensibilité de cette corrélation, dans un premier temps on propose de calcu-

ler l'incertitude des scores d'impact ACV en faisant l'hypothèse de corrélation « **intradonnées** », c'est-à-dire de la manière habituelle. Dans un second temps on calcule l'incertitude des scores d'impact ACV sans faire l'hypothèse de corrélation intradonnées qui revient à corrélé les différentes apparitions d'un même processus dans le cycle de vie. Les résultats des incertitudes obtenus dans les deux cas sont finalement comparés afin de connaître la sensibilité de cette hypothèse.

4.1.1 Propagation de l'incertitude dans le cas de l'hypothèse de corrélation intradonnées totalement corrélée

Le cas totalement corrélé correspond à la situation classique des logiciels ACV. C'est-à-dire en considérant l'arbre infini comme son réseau associé comme on peut le voir dans la Figure 4.1. Afin d'étudier la propagation de l'incertitude des impacts dans cette situation, plusieurs étapes vont être appliquées dans un programme développé sous le langage de programmation **Python**.

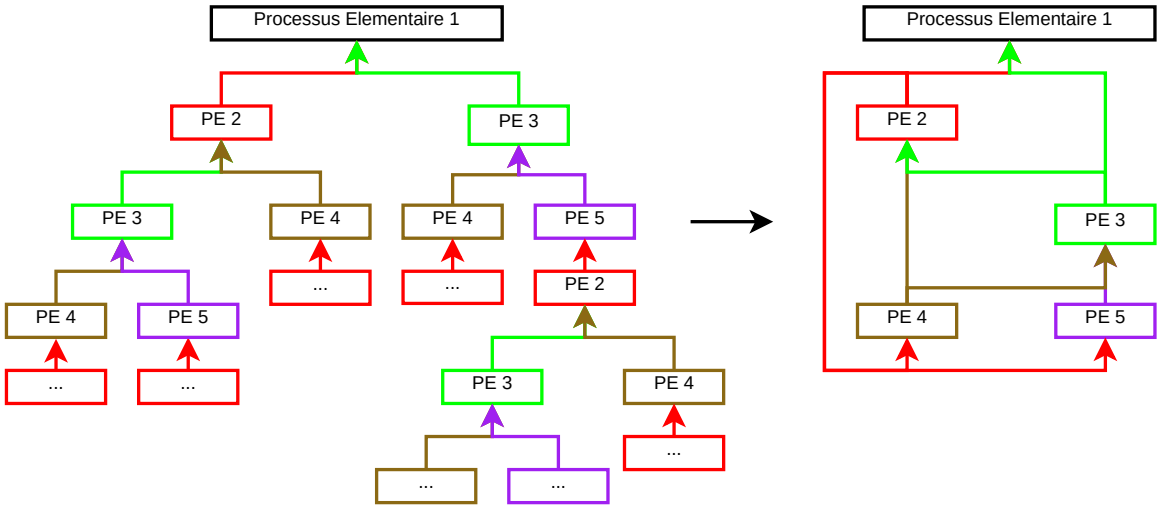


Figure 4.1 Passage de l'arbre infini (réalité) au réseau (hypothèse de corrélation **intradonnées**).

La première étape est de déterminer la matrice technologique **A** associée au réseau complet des processus de la base de données **ecoinvent v2.2** à partir d'un fichier d'exportation du logiciel **SimaPro**. Cette base contient 4161 processus élémentaires, la matrice technologique est ici de la taille 4161x4161. Elle est creuse (43 616 valeurs non nulles, soit $0,2\%$ des valeurs), avec une diagonale composée quasi systématiquement de 1, le reste des valeurs sont quasi systématiquement négatives. La majorité les éléments de la matrice **A** sont des variables aléatoires (89.58%) qui suivent des lois log-normales (89.5%), normales (0.06%) ou

triangulaires (0.02%). Leurs distributions de probabilité sont stockées dans des matrices. Cela permet de stocker seulement des données nécessaires de la matrice creuse. Quatre matrices sont créées : une pour chaque loi plus une pour les valeurs déterministes non nulles. Ces matrices contiennent autant de lignes qu'il y a variable aléatoire dans la distribution associée. Les deux premières colonnes de la matrice contiennent les coordonnées de la variable (la ligne et la colonne dans la matrice $\mathbf{A}$), puis les colonnes suivantes contiennent les paramètres de la loi (seulement la valeur déterministe est considérée dans le cas déterministe).

La deuxième étape est de déterminer la matrice intervention $\mathbf{B}$ qui contient les flux élémentaires de chacun des processus. Tout comme la matrice technologique déterminée précédemment, les éléments de la matrice $\mathbf{B}$ sont majoritairement des variables aléatoires (60.15%). Cette matrice contient 4161 colonnes (une par processus élémentaire) et 1615 lignes (une par flux élémentaire). Cette matrice aussi est creuse (93 087 valeurs non nulles, soit $1,4\%$), car un processus n'a que quelques flux élémentaires directs parmi tous ceux répertoriés dans la base de données. Les variables aléatoires qui composent cette matrice suivent majoritairement des lois log-normales (60.13%) ainsi que quelques lois normales (0.02%). La même méthode est appliquée pour stocker les variables de cette matrice.

La matrice des facteurs caractéristiques $\mathbf{C_F}$ est considérée comme déterministe dans notre étude. Ses valeurs sont connues et proviennent d'un fichier d'exportation du logiciel **SimaPro**. Cette matrice compte 1615 lignes et 5 colonnes, car elle permet de catégoriser les 1615 flux élémentaires dans 5 catégories de dommages.

Une simulation de Monte-Carlo permet de déterminer les distributions estimées des résultats. Le calcul des impacts est le suivant :

$$\mathbf{H} = \mathbf{C_F^\top B A^{-1}}$$

Dans cette équation, la matrice identité est utilisée comme vecteur de demande ici. Ainsi le résultat n'est pas un vecteur des impacts associé à un vecteur de demande, mais une matrice contenant dans chaque colonne les impacts associés à une unité de production d'un processus élémentaire. Plus en détail, à chaque itération, les matrices $\mathbf{A}$ et $\mathbf{B}$ sont initialisées à la matrice nulle aux bonnes dimensions puis les variables aléatoires sont tirées et enfin placées grâce aux informations d'incertitudes stockées. Dans le cas de l'utilisation d'un ordinateur de bureau assez performant avec 16Go de RAM, ces étapes sont très rapides (~ 0.1 seconde). Ensuite, le calcul est appliqué et celui-ci est plus long, car la matrice $\mathbf{A}$ est relativement grande et doit être inversée (~ 12 secondes) puis multipliée à la matrice $\mathbf{B}$ qui est aussi relativement grande (~ 3 secondes). Cette étape de calcul dure environ 15 secondes par itérations. Ce qui

donne pour une simulation de 10 000 itérations 150 000 secondes, soit presque 2 jours.

Finalement, le résultat obtenu est une hyper matrice dans laquelle une dimension correspond aux 5 catégories de dommages associés aux résultats d'impact, une dimension contient les 4161 produits unitaires associés aux processus élémentaires et une dimension contient 10 000 itérations de chaque variable.

Une étude statistique est alors réalisée sur ces échantillons pour déterminer l'allure des distributions ainsi que l'incertitude des résultats obtenus grâce aux estimations de la variance et de la moyenne, ou plus exactement, au coefficient de variation. En annexe B une étude montre que les distributions des impacts au niveau du dommage convergent vers des lois Log-normales dont les paramètres ont été approximés pour tous les processus grâce aux échantillons des simulations de Monte-Carlo.

4.1.2 Propagation de l'incertitude dans le cas décorrélé

Comme il a été énoncé plus haut, l'objectif à présent est de calculer l'incertitude des impacts sans faire l'hypothèse de corrélation intradonnées. Pour cela, une solution est de passer par un arbre tronqué dans lequel il n'y aura pas de rétroactions. Un outil développé sous le langage de programmation **Python** permet la construction de ces arbres correspondants à des systèmes de produits désagrégés (Bourgault et al., 2012).

La première étape est alors de construire l'arbre tronqué associé au système de processus que l'on souhaite étudier. Cette méthode consiste à partir de l'arbre de processus représentant la réalité, puis d'arrêter l'arbre lorsque les impacts des nœuds (processus) sont devenus négligeables devant l'impact du système total. Cependant, « arrêter » ici ne signifie pas que les impacts ne sont pas considérés, mais que le processus d'arrêt est considéré comme agrégé. C'est-à-dire que l'arborescence de l'arbre est arrêtée et les flux élémentaires associés au cycle de vie totale de la demande de ce processus sont considérés. Cela revient à dé-corréler tous les processus les plus impactant dans le système de produit. Mais lorsque la contribution aux impacts devient inférieure à un seuil prédéterminé, les données agrégées sont utilisées. Le cas décorrélé correspond donc en réalité à une approximation d'une situation dans laquelle tous les processus seraient décorrélés. Cependant l'arbre serait infini et le calcul impossible.

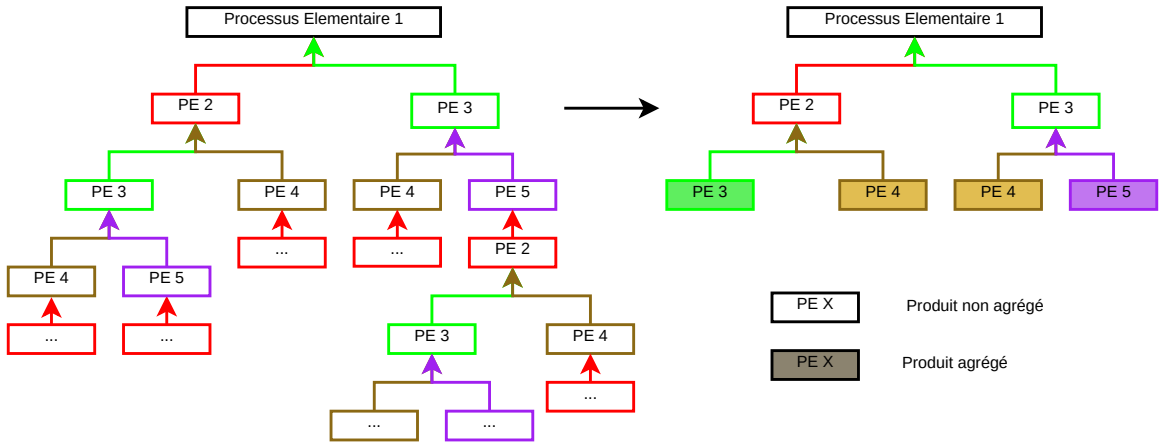


Figure 4.2 Passage de l'arbre infini (réalité) à l'arbre tronqué (sans l'hypothèse de corrélation **intradonnées**).

Dans la Figure 4.2, il est possible de voir comment passer d'un arbre infini à un arbre tronqué. Les nœuds pleins correspondent à des processus élémentaires dont les impacts seront agrégés. C'est-à-dire que l'impact de ce processus et de tous ceux qui le suivent est inférieur à un certain pourcentage de l'impact total du système. Le critère à partir duquel il faut agréger les impacts des processus élémentaires et donc stopper la progression de l'arbre est déterminé arbitrairement selon la précision souhaitée pour la désagrégation. Par exemple, pour un critère de désagrégation de 1%, si les impacts agrégés d'un nœud sont inférieurs à 1% des impacts du système total dans toutes les catégories de dommage, alors les impacts de ce nœud sont agrégés et il devient une feuille de l'arbre.

Une fois l'arbre tronqué construit, le graphe obtenu n'est plus infini, il est alors possible de construire la nouvelle matrice technologique $\mathbf{A}$ associée à la structure de ce graphe comme il est présenté dans la Figure 4.3. Dans chaque colonne (correspondant à un nœud) est indiquée la quantité des autres nœuds nécessaires à sa production, et un 1 sur la diagonale, car la production d'un processus produit une unité du produit associé. Ainsi, lorsque le processus est agrégé, il n'y a qu'un 1 sur la diagonale de la colonne.

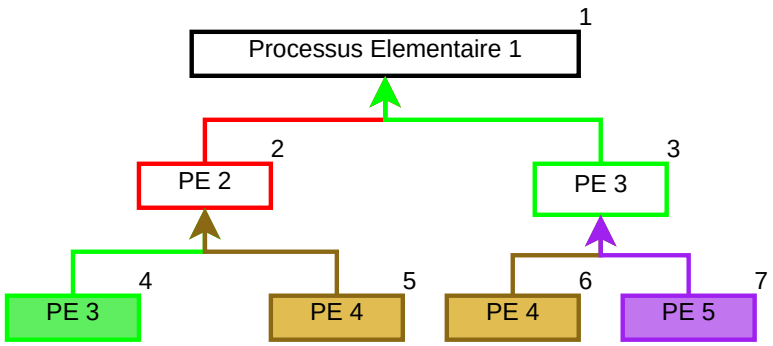


Figure 4.3 Construction de la matrice technologique à partir de l'arbre tronqué.

$$A = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & . & . & . & . \\ * & 1 & 0 & . & . & . & . \\ * & 0 & 1 & . & . & . & . \\ 0 & * & 0 & 1 & . & . & . \\ 0 & * & 0 & . & 1 & . & . \\ 0 & 0 & * & . & . & 1 & . \\ 0 & 0 & * & . & . & . & 1 \end{pmatrix} \end{array} \quad \textit{Flux économiques}$$

$$B = \begin{pmatrix} 0 & * & 0 & . & . & . & . \\ * & 0 & 0 & . & . & . & . \\ * & 0 & * & . & . & . & . \\ 0 & * & 0 & . & . & . & . \\ \hline . & . & . & * & * & * & * \\ . & . & . & * & * & * & * \\ . & . & . & * & * & * & * \end{pmatrix} \begin{array}{l} \textit{Flux élémentaires} \\ \\ \\ \\ \\ \textit{Impacts agrégés} \\ \\ \end{array}$$

Ensuite la construction de la matrice intervention associée se fait comme suit : lorsque le processus n'est pas agrégé, ses émissions directes dans la nature sont considérées, c'est-à-dire la colonne de la matrice intervention complète associée à ce processus. Cependant si le processus a été agrégé, il faut considérer les émissions associées à la totalité de son cycle de vie. Autrement dit, les éléments de la matrice interventions de la colonne d'un processus agrégé correspondront aux éléments de la colonne de ce processus dans la matrice $\mathbf{BA^{-1}}$ avec les matrices $\mathbf{A}$ et $\mathbf{B}$ d'origine.

Il est important de rappeler que les éléments qui composent les nouvelles matrices technolo-

giques sont des variables aléatoires. Donc lorsque celles-ci proviennent de la matrice technologique d'origine ou de la matrice d'intervention d'origine, ces informations sont disponibles directement dans la base de données **ecoinvent v2.2**. Cependant, pour les émissions agrégées de la nouvelle matrice intervention, les éléments proviennent de la matrice intensité $\mathbf{BA^{-1}}$ et les informations d'incertitudes ne sont pas directement disponibles. C'est pourquoi elles ont été précédemment calculées par la méthode de Monte-Carlo. En réalité, pour des questions de gain de temps, ce sont les informations d'incertitudes qui ont été déterminées dans la section 4.1.1 qui ont été utilisées. Cependant, ces informations n'étaient disponibles que pour les impacts au niveau dommage et pas pour les émissions directement (après multiplication par la matrice $\mathbf{C_F}$). Quelques transformations ont alors été appliquées à la nouvelle matrice intervention. Cette matrice contient une ligne pour chaque émission, plus 5 lignes qui correspondent à chacune des catégories de dommages. Pour une colonne d'un processus qui est agrégé dans l'arbre, toutes les émissions sont mises à 0 et ce sont directement des informations de dommages dans les 5 lignes supplémentaires qui sont indiquées. Il est important de rappeler ici que les informations d'incertitude renseignées sont les paramètres correspondant à la loi log-normale approximant la distribution des impacts (la méthode des moments qui se base sur l'esperence mathématiques et la variance mathématique est utilisée). Cette méthode estime les paramètres de la loi log-normale en conservant les deux premiers moments issus de l'échantillon.

Une fois les nouvelles matrices (technologique et intervention) construites, comme précédemment une simulation de Monte-Carlo est appliquée sur l'équation suivante :

$$\mathbf{h} = \mathbf{C_F^\top B A^{-1} f}$$

Cependant, il faut faire attention ici, car aucun des éléments n'est le même que précédemment (section 4.1.1). Le vecteur $\mathbf{f}$ ici est un vecteur composé uniquement de 0 avec seulement un 1 en première position. La matrice $\mathbf{A}$ est la nouvelle matrice technologique déterminée, elle contient autant de ligne et de colonne qu'il y a de nœud dans l'arbre. Des processus similaires apparaissant à plusieurs positions dans l'arbre correspondent donc à des colonnes différentes dans les matrices $\mathbf{A}$ et $\mathbf{B}$. La matrice $\mathbf{B}$ est la nouvelle matrice intervention, elle contient autant de colonnes qu'il n'y a de nœud dans l'arbre et autant de lignes qu'il y a d'émissions considérées plus 5 lignes supplémentaires correspondant aux impacts au niveau des catégories de dommage. Finalement, la matrice $\mathbf{C_F}$ des facteurs de caractérisation ne peut être la même que précédemment, car le nombre de lignes de la matrice intervention a changé. La matrice $\mathbf{C_F}$ contient alors 5 colonnes en plus (correspondant aux 5 catégories de dommages ajoutées). Les 1615 premières colonnes sont les mêmes que dans la partie précédente, et les 5 dernières

colonnes forment une matrice 5x5 identités. Finalement, le résultat obtenu ici (**h**) correspond au vecteur des impacts dans les 5 catégories de dommages pour le processus observé.

Il est important de remarquer ici que la matrice technologique et la matrice intervention sont spécifiques à chaque processus observé, car chaque arbre est différent. C'est pourquoi les calculs pour cette étape sont plus conséquents. Comme précédemment, l'étape de calcul qui prend le plus de temps dans la simulation de Monte-Carlo est l'inversion de la matrice **A** et le produit avec la matrice **B**. La taille de ces matrices peut être directement contrôlée par le critère de désagrégation au détriment de la précision de cette désagrégation. Un critère de 1% permet d'avoir un bon compromis entre la précision et la taille des matrices ($\sim$ 500 processus dans l'arbre). Cela donne un temps d'inversion de la matrice technologique d'environ 4 secondes. Une itération de la simulation de Monte-Carlo dure environ 1 seconde. Les simulations de Monte-Carlo sont faites ici avec 1000 itérations pour des raisons de temps. La plus grande différence avec le cas précédent est qu'il faut faire une simulation pour chaque système unitaire de processus, c'est-à-dire pour chaque processus, soit 4161 simulations. Cette étape est celle qui prend le plus de temps de calcul dans l'étude avec environ 50 jours si tous les calculs sont faits sur une seule machine. Mais ce temps peut être divisé avec du calcul parallèle.

Ensuite, comme dans la partie précédente, les informations statistiques sont tirées des échantillons obtenus à la suite des simulations et sont comparées à celles du cas totalement corrélé.

4.1.3 Comparaison des deux incertitudes

La comparaison des informations d'incertitudes entre le cas totalement corrélé et le cas partiellement corrélé se fait à travers le rapport des coefficients de variation. Le coefficient de variation est choisi ici, car il est sans dimension. Ainsi, pour chaque processus de la base de données, dans chacune des catégories de dommage, le rapport entre les coefficients de variation des deux échantillons obtenus est observé. Étant donné le nombre important de processus à observer dans la base de données **ecoinvent v2.2**, des box-plot sont réalisés afin d'avoir une vue d'ensemble des résultats. Ensuite, les cas de forte différence seront observés plus en détail.

$$\alpha_i = \frac{C_v(h_i^{\text{cas corrélé}})}{C_v(h_i^{\text{cas non corrélé}})}$$

$\forall\ h_i$: impact du processus élémentaire i pour une des catégories de dommages

Si l'on observe que le rapport entre les coefficients de variation est égal à 1, alors cela signifie

que l'hypothèse de corrélation **intradonnées** n'est pas sensible dans la propagation des incertitudes en ACV. Dans le cas où le rapport α est supérieur à 1, cela signifie que les résultats d'impacts obtenus avec l'hypothèse de corrélation sont les plus incertains.

4.1.4 Étude des processus responsables de la sensibilité

Une fois la sensibilité à l'hypothèse de corrélation déterminée, il est intéressant de savoir si certains processus sont responsables de cette sensibilité. Pour cela, l'arborescence des arbres est analysée et un score de désagrégation **S** est calculé pour chaque flux économique. La corrélation entre ce score et la sensibilité de l'hypothèse de corrélation permet d'identifier les flux économiques susceptibles d'être responsables de cette sensibilité.

Le score de désagrégation est déterminé en fonction de deux variables : le nombre d'occurrences d'un flux économique i dans l'arbre tronqué du système du processus élémentaire j (noté n_{ij}), et le pourcentage de sa contribution à l'impact du système complet (noté q_{ij}). Ces scores de désagrégations sont stockés dans une matrice de la taille de la matrice technologique : autant de colonne qu'il y a de processus élémentaire (arbre tronqué) et autant de ligne qu'il y a de flux économique. Chaque colonne correspond au système du processus élémentaire de l'arbre observé et chaque ligne correspond au flux élémentaire associé à un processus élémentaire de la base de données **ecoinvent v2.2**. L'élément de la colonne j et de la ligne i correspond donc au score noté s_{ij} calculé comme suite :

$$s_{ij} = \max(0, n_{ij} - 1) q_{ij}$$

$$\mathbf{S} = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{l_1} \\ \vdots \\ \mathbf{l_n} \end{pmatrix} \quad \begin{cases} \forall \text{ flux économique associé au processus élémentaire } i \\ \forall \text{ arbre tronqué associé au processus élémentaire } j \end{cases}$$

Dans cette équation, n_{ij} correspond au nombre d'occurrences du processus i dans l'arbre tronqué du système du processus j, si n égale 0 ou 1, le processus n'est pas désagrégé. Son score est alors nul. Ensuite, si le processus est désagrégé (occurrences multiples dans l'arbre tronqué), alors le score de désagrégation est proportionnel au nombre d'occurrences dans l'arbre tronqué et au pourcentage de son impact dans le système.

Chaque ligne de la matrice S correspond aux différents scores d'un même flux économique pour chacun des arbres tronqués, ces lignes sont notées $\mathbf{l_i}$. L'information cherchée est alors la corrélation entre les éléments d'une ligne $\mathbf{l_i}$, soit les scores de désagrégations d'un flux économique pour chaque arbre, et les indicateurs de sensibilité de chacun de ces arbres. Ces indicateurs de sensibilité correspondent au rapport des coefficients de variations dans les cas

totalement corrélé et décorrélé. Ces indicateurs sont stockés dans un vecteur noté α qui a autant d'éléments que de processus élémentaire dans la base de données **ecoinvent v2.2**.

$$C_i = \mathrm{Cor}(\mathbf{l_i}, \alpha) \qquad \forall \text{ flux écomomique associé au processus élémentaire } i$$

La matrice des scores étant déterminée, pour chaque flux économique i correspondant à un processus de la base de données (chaque ligne $\mathbf{l_i}$), la corrélation entre les scores de désagrégation de ce flux économique dans tous les arbres et la sensibilité de tous les arbres (rapport des coefficients de variations) est observée. Ainsi la corrélation entre la désagrégation d'un flux économique et la sensibilité à l'hypothèse de corrélation est déterminée et il est possible de connaître les flux économiques associés à la sensibilité de l'hypothèse.

4.2 Corrélation intersystème dans l'incertitude d'un système de produits construits avec des processus agrégés

L'étude de la partie précédente a donc permis de comparer deux méthodes de calcul de la propagation des incertitudes pour des systèmes unitaires de la base de données **ecoinvent v2.2**.

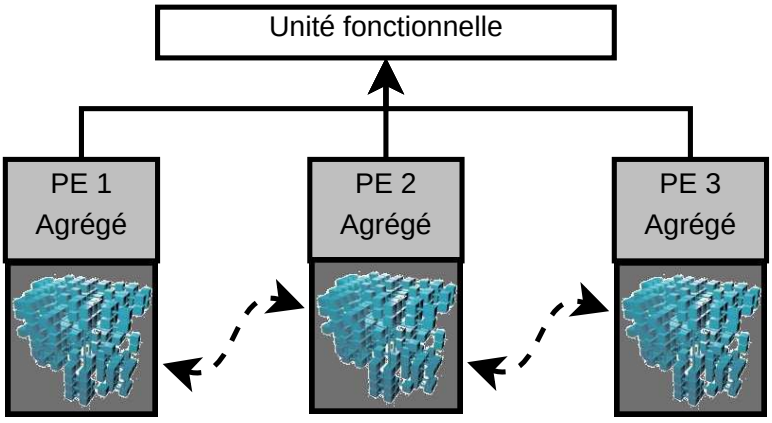


Figure 4.4 Corrélation intrasystème.

Les simulations de Monte-Carlo pour chacun des processus de la base de données ont permis d'obtenir des distributions empiriques des impacts agrégés des processus élémentaires. Ces distributions empiriques sont approximées par des distributions Log-normales grâce à la méthode des moments. Cela permet de ne stocker que les paramètres μ et σ de chaque distribution approximée. Ces deux paramètres suffisent à déterminer la valeur déterministe de l'impact (mode, moyenne ou médiane) avec une donnée d'incertitude (variance) grâce aux

équations suivantes.

$$X \sim LN(\mu, \sigma) \begin{cases} \text{mode}(X) = e^{\mu - \sigma^2} \\ \text{E}(X) = e^{\mu + \frac{\sigma^2}{2}} \\ \text{Var}(X) = e^{2(\mu + \sigma^2)}(1 - e^{-\sigma^2}) \end{cases}$$

Les données d'incertitudes étant à présent connues pour les impacts agrégés des processus élémentaires de la base de données agrégée **ecoinvent v2.2**, il reste maintenant à déterminer l'incertitude d'un système de produit qui est composé de ces processus élémentaires.

Un système de produit composé de processus agrégé correspond à une combinaison linéaire de ces processus pondérés de leurs quantités. Le calcul de l'impact est linéaire et correspond donc à la somme des impacts de chaque processus pondéré de la quantité du processus nécessaire. Cependant en ce qui concerne l'incertitude, le calcul de la variance n'est pas linéaire et fait intervenir les covariances entre ces données agrégées. Les covariances ne sont pas nulles, car le calcul des impacts de deux processus élémentaires différents fait appel aux mêmes flux élémentaires et flux économiques.

4.2.1 Matrices de variances-covariances de tous les processus élémentaires

Il faudra construire une matrice de variance-covariance des impacts des processus dans chacune des catégories de dommage considérées.

Les incertitudes des impacts des processus agrégés, dépendent dans chacune des deux méthodes utilisées des mêmes variables aléatoires, celles des matrices technologiques et matrices interventions. C'est pourquoi il est certain que leurs impacts ne sont pas indépendants. L'utilisation des simulations de Monte-Carlo permet d'estimer les matrices de variances-covariances. Cependant, les simulations de la méthode avec l'utilisation des arbres tronqués ne permettent pas cette estimation, car pour chaque processus, de nouveaux tirages indépendants sont réalisés. Cela provient du fait que les matrices (technologique et intervention) sont spécifiques au processus analysé. Cependant dans le cas de l'hypothèse de corrélation totale, le fait d'utiliser la matrice identité comme demande au lieu d'un vecteur de demande classique, permet d'avoir les mêmes éléments des matrices technologiques et intervention tirée pour tous les processus agrégés lors d'une itération de la simulation de Monte-Carlo. Ainsi, les matrices de variances-covariances sont construites à partir des échantillons de la simulation de Monte-Carlo du cas avec l'hypothèse de corrélations intradonnées totales.

4.2.2 Calcul de l'incertitude d'un système de produits agrégés non indépendants

Une fois les matrices de variances-covariances estimées, il est possible de calculer l'incertitude et plus précisément la variance d'un système de produit constitué d'une combinaison linéaire de processus agrégés. L'équation de la variance d'une combinaison linéaire suivante est utilisée :

$$\mathrm{Var}(\sum_{i=1}^{n} a_i h_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(h_i, h_j)$$

Dans cette équation, les h_i représentent les impacts des processus agrégés dans une catégorie de dommage et les a_i les quantités associées dans le système. Cette formule permet alors de déterminer la variance d'un système tout en considérant les covariances entre les impacts des différents processus agrégés qui le composent.

4.2.3 Part des covariances dans l'incertitude du système total

Dans l'équation précédente, il est possible de séparer le calcul de la variance en deux termes distincts. Un premier terme est composé des variances des différents impacts des processus, cela correspond au cas où i=j dans la double somme. Un deuxième terme composé des covariances entre les impacts des processus, cela correspond à tous les autres cas, lorsque $i \neq j$.

$$r = \frac{\sum_{i=1}^{n}\sum_{j=1\ j\neq i}^{n} a_i a_j \mathrm{Cov}(h_i, h_j)}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(h_i, h_j)}$$

L'évaluation de la part des covariances dans la variance totale d'un système est calculée avec un système proposé dans une étude de cas. Cette étude de cas correspond à un système de produit composant le cycle de vie d'un burger. Les étapes de production, d'emballage et de transports sont considérées pour un total de 16 processus agrégés qui ne sont pas explicités pour des raisons de confidentialité.

4.3 Corrélation intersystème dans la comparaison de systèmes de produits construits avec des processus agrégés

La corrélation entre les différents processus agrégés qui composent un système de produit est aussi présente lors de la comparaison de deux systèmes différents. Cette corrélation est du même ordre que cette de la partie précédente, elle correspond à celle entre les processus qui

composent les systèmes. Cependant, elle s'applique lors de la comparaison de deux systèmes.

4.3.1 Comparaison de deux systèmes en considérant leurs corrélations

Lorsque les impacts de deux systèmes de produits différents sont comparés, il peut être intéressant de tenir compte de l'incertitude des résultats. De plus, lors de la prise en compte des incertitudes associées aux résultats il est important de considérer les corrélations entre les différents processus présents dans les deux systèmes.

Afin de tenir compte de ces incertitudes dans la comparaison, un nouveau système est considéré. Ce système correspond à la différence des deux systèmes que l'on observe. Ainsi, tout comme dans la partie précédente, une combinaison linéaire de processus agrégés est analysée. De plus, une hypothèse sur la normalité de la différence des impacts de deux systèmes est posée. Cette hypothèse a été étudiée dans l'annexe C, il est possible d'approximer la distribution avec la méthode des moments en estimant les paramètres de la loi. En effet, les paramètres de la loi normale peuvent être estimés avec la valeur déterministe correspond à la moyenne, et l'écart type obtenu avec le calcul de la variance en considérant toutes les covariances. Finalement pour comparer les deux systèmes de produits, il suffit d'observer la position de la distribution autour de 0. La probabilité que la différence soit supérieure à 0 correspond à la probabilité que le premier système soit moins impactant et vice versa.

4.3.2 Permettre à l'utilisateur de spécifier l'indépendance des mêmes produits utilisés sans les systèmes.

Il a été jugé intéressant que les utilisateurs des logiciels ACV qui construisent le système de produit puissent spécifier si les différentes apparitions d'un même processus dans les systèmes sont indépendantes. Cette indépendance provient par exemple de l'utilisation d'une source différente pour un ingrédient par exemple, ou l'utilisation de deux camions différents pour la livraison des deux produits. Cette spécification apparaît lors de la construction des systèmes de produits. L'utilisateur a la possibilité de cocher une case qui précise que les processus ne sont pas corrélés. Dans le calcul, cela revient à considérer la covariance entre les deux processus nulle.

4.3.3 Fournir une information pertinente au meilleur choix de système

Le résultat fourni à l'utilisateur par le logiciel doit être clair. Ce résultat apparaît alors sous la forme d'une barre avec d'un côté une couleur correspondant au cas où le système 1 et préférable au système 2, et de l'autre côté le cas contraire. Ensuite, plus la probabilité

d'un cas est grande, plus sa partie correspondante de la barre est grande. De cette manière, l'utilisateur peut connaître directement la probabilité qu'un système soit préférable à un autre.

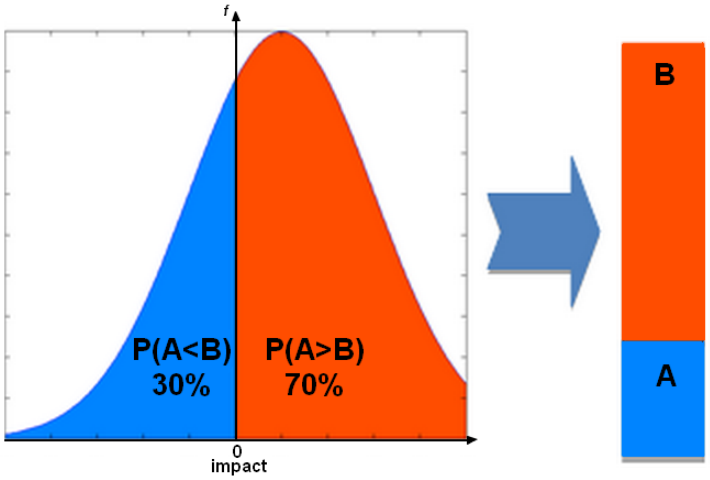


Figure 4.5 Représentation graphique lors d'une comparaison de systèmes de produits.

4.4 Intégration de la prise en compte des incertitudes dans les logiciels d'éco-design.

Les parties précédentes permettent donc de calculer les impacts d'un système de produit construit à partir de processus agrégés et d'y calculer l'incertitude des résultats. Il est possible de choisir une base de données avec ou sans l'hypothèse de corrélation intradonnées. Les corrélations entre les différents processus d'un système sont considérées ainsi que la corrélation des processus lors de la comparaison de deux systèmes.

4.4.1 Développement d'un outil

Afin de tester et de mettre en application les résultats de l'étude, un outil de type « Preuve de concept » a été construit. Cet outil est développé sous le langage de programmation **Python**. Il doit respecter les contraintes des logiciels ACV simplifiés utilisés par les industriels (simplicité d'utilisation, rapidité d'exécution, clarté des résultats . . .) mais doit aussi informer des incertitudes en prenant en compte les différents niveaux de corrélation de l'étude.

Calcul des impacts avec incertitudes

L'outil permet de construire deux systèmes de produits. Les processus disponibles sont les données agrégées de la base de données **ecoinvent v2.2**. L'outil somme les impacts des différents processus composant les systèmes en les pondérant de leurs quantités pour obtenir l'impact du système complet. En ce qui concerne les incertitudes, la variance est calculée en fonction des variances des systèmes agrégés et de leurs covariances. Finalement, les impacts et les incertitudes sont représentés sur des graphes simples. De plus, le graphique permettant la comparaison des deux systèmes présentés dans la partie précédente est disponible.

Prise en compte des différents niveaux de corrélations dans le calcul

Les différents niveaux de corrélations vues dans l'étude sont pris en compte dans l'outil. Premièrement, il est possible de choisir si on souhaite utiliser les incertitudes des processus agrégés avec ou sans l'hypothèse de corrélation. Il est de plus possible de faire plusieurs fois le calcul en faisant varier le critère de désagrégation pour ensuite choisir les incertitudes désirées dans l'outil. Celui-ci fait appel à différents jeux de données. Les paramètres de la loi Log-normale de chaque processus agrégé sont conservés pour chaque catégorie de dommage et chaque critère de désagrégation.

Deuxièmement, lors du calcul de la variance du système complet, la covariance entre les différents processus qui composent les systèmes est considérée. Ce niveau de corrélation est aussi considéré lors de la comparaison de deux systèmes. De plus, comme il est mentionné dans la section 5.3.2, l'utilisateur peut choisir de rendre indépendants deux processus similaires apparaissant dans les deux systèmes.

Réponds aux exigences de rapidités de l'utilisateur

Les incertitudes ont été pré-calculées par les simulations de Monte-Carlo et seulement les paramètres des lois de probabilités Log-normale qui approximent les résultats sont conservés. Cela permet alors d'accéder directement aux incertitudes des impacts des processus élémentaires sans avoir à refaire de simulation de Monte-Carlo à chaque fois. De plus, les matrices de variances-covariances sont chargées et stockées lors de l'ouverture du logiciel. Cela permet donc d'accéder directement aux covariances lors du calcul. Les matrices de variances-covariances sont très grandes, mais contiennent beaucoup de valeurs doublées. En effet, la symétrie de la matrice implique que seulement la moitié des données est nécessaire à stocker. De plus, la diagonale correspond aux variances qui sont accessibles grâce aux paramètres des lois déjà stockées. Ainsi seulement les données nécessaires sont chargées et stockées pour

optimiser la mémoire utilisée et le temps de chargement.

De plus, l'utilisation de calcul parallèle donne la possibilité de charger les données nécessaires au calcul pendant que l'outil est actif. Cela permet à l'outil de charger les données pendant que l'utilisateur construit les systèmes de produits. Ainsi le temps de chargement n'est pas perçu par l'utilisateur. Finalement, lorsque le calcul des impacts est lancé, les données ont déjà été chargées et le calcul qui est une combinaison linéaire des impacts des différents processus pondérés de leurs quantités peut être effectué instantanément.

Facile à intégrer dans le logiciel d'écodesign

La réalisation de l'outil développé avec python nous permet d'appliquer l'étude. De plus, cela permet d'avoir un exemple de code d'application sous le langage Python. Les calculs réalisés n'ont rien de spécifique au langage de programmation et peuvent être répliqué sous d'autre langages de programmation comme **Java**, **PHP** ou encore **C++**. Ainsi, le service informatique du partenaire industriel pourra aisément intégrer l'étude au logiciel qu'ils utilisent en adaptant celui de l'outil.

4.4.2 Application à l'étude de cas proposée par le client.

Afin de tester l'étude réalisée, une étude de cas a été fournie par un partenaire industriel. Cette étude de cas comprend l'étude des impacts de deux systèmes de produits détaillés dans le Tableau 4.1 ainsi que leur comparaison. Dans cette étude de cas, les incertitudes des résultats d'impact sont analysées en considérant les différents niveaux de corrélations de l'étude. L'outil est testé sur cette étude de cas, afin de vérifier les différents résultats et les différentes fonctionnalités qu'il propose.

Tableau 4.1 Valeurs des facteurs de la matrice pedigree selon leur note.

Processus ecoinvent	**Veggie burger** (x10^{-3})	**Meat Burger** (x10^{-3})	**Unit**
Grass silage organic, at farm/CH U	0	4 630	kg
Tap water, at user/RER U	40 000	4 000	kg
Soybeans, at farm/US U	150	0	kg
Wheat IP, at feed mill/CH U	6.82	113.7	kg
Grass silage organic, at farm/CH U	170	110	kg
Electricity, production mix RER/RER U	60	72	kWh
Electricity, production mix RER/RER U	30	30	kWh
Natural gas, high pressure, at consumer/RER U	225	270	MJ
Tap water, at user/RER U	450	540	kg
Fleece, polyethylene, at plant/RER U	5.70	0	kg
Packaging film, LDPE, at plant/RER U	3.10	0	kg
Packaging, corrugated board, mixed fibre, single wall, at plant/RER U	0	22.2	kg
Packaging, corrugated board, mixed fibre, single wall, at plant/RER U	30	0	kg
Packaging film, LDPE, at plant/RER U	1.13	0.00	kg
Packaging, corrugated board, mixed fibre, single wall, at plant/RER U	0	14.5	kg
Packaging film, LDPE, at plant/RER U	0.125	0.162	kg
EUR-flat pallet/RER U	0.0168	0.0092	p
Transport, lorry 7.5-16t, EURO5/RER U	22.5	27.0	tkm
Transport, lorry 7.5-16t, EURO5/RER U	22.5	27.0	tkm

CHAPITRE 5 RÉSULTATS

Dans ce chapitre, les résultats de l'étude réalisée sont présentés. L'application de la méthodologie du chapitre 4 est appliquée. Les résultats sont généralement présentés pour chacune des cinq catégories de dommage de la méthode d'impact utilisée.

5.1 Sensibilité de l'hypothèse de corrélation intradonnées

Afin de présenter clairement la sensibilité de l'hypothèse de corrélation intradonnées, il est en premier lieu présentés les informations de l'incertitude dans le cas corrélé puis dans le cas décorrélé. Ensuite, une analyse sur rapport des coefficients de variation dans les deux cas est faite. Pour finir, les sources potentielles de cette sensibilité sont exposées.

5.1.1 Propagation de l'incertitude dans le cas totalement corrélé

Dans un premier temps, la propagation de l'incertitude dans le cas totalement corrélé est analysée. Pour rappel, ce cas correspond à une corrélation totale des processus identiques apparaissant à différentes reprises dans l'arbre du système de produit. Cette hypothèse est faite lors du passage de l'arbre infini à un réseau, ou encore lors du passage de la série entière à la matrice technologique inverse.

Une simulation de Monte-Carlo est réalisée et donne une distribution empirique des impacts pour chacune des catégories de dommage. Un exemple de résultat est donné dans la série des graphes de la Figure 5.1. Cet exemple correspond au processus « *Electricity, hard coal, at power plant/UCTE U* » qui est un processus de la base de données utilisée (**ecoinvent**) et choisi arbitrairement.

La courbe de tendance est obtenue tout d'abord en faisant l'hypothèse que la distribution obtenue pour un impact au niveau dommage est log-normale, puis en estimant les paramètres de la loi Log-normale avec les 1er et 2^{e} moments de l'échantillon obtenu après la simulation de Monte-Carlo. Cela correspond à l'approximation de la loi par la méthode des moments. À l'issue de la simulation de Monte-Carlo, la variance et la moyenne de l'échantillon sont calculés. Puis les paramètres μ et σ de la loi log-normale correspondants sont estimés à partir des équations suivantes :

$$\hat{\mu} = \ln\left(\mathrm{E}(X)\right) - \frac{1}{2}\ln\left(1 + \frac{\mathrm{Var}(X)}{\mathrm{E}(X)^2}\right)$$

$$\hat{\sigma^2} = \ln\left(1 + \frac{\mathrm{Var}(X)}{\mathrm{E}(X)^2}\right)$$

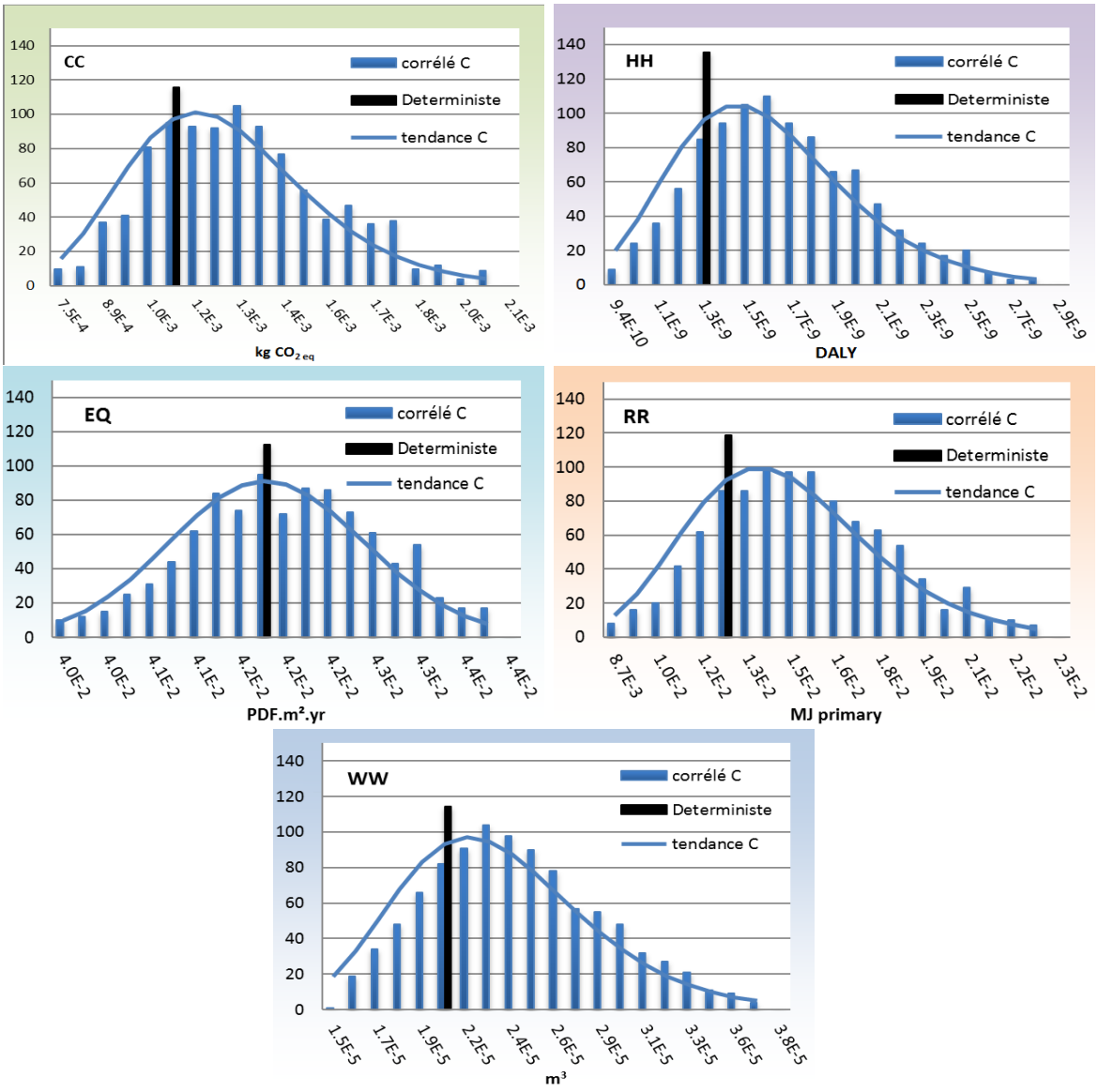


Figure 5.1 histogrammes (1000 tirages) et approximation log-normale (tendance) des impacts de la production de 1 kWh de « *Electricity, hard coal, at power plant/UCTE U* ».

Une étude a été faite pour vérifier l'hypothèse de log-normalité, cette étude est présente en annexe B et nous permet de conclure que la distribution des impacts au niveau du dommage peut être approximée par une loi Log-normales.

Ce résultat est obtenu pour l'intégralité des systèmes de produits pouvant être construits

avec la base de données **ecoinvent v2.2**. Cette base de données contenant plus de 4000 processus, une étude statistique a été réalisée sur l'ensemble de ces processus. Une attention particulière est portée sur deux variables. Une première variable est le rapport de la médiane de la tendance sur la valeur déterministe. La valeur déterministe correspond à l'impact obtenu lorsque les valeurs déterministes des variables aléatoires sont prises dans les matrices **A** et **B** lors du calcul de l'impact présenté dans la section 4.1.1. Il est effectivement intéressant de savoir où se situe la valeur obtenue de manière déterministe comparée à la médiane de la distribution. Car il est attendu que ces valeurs soient confondues. La deuxième variable analysée est le coefficient de variation qui correspond au rapport entre l'écart type et la moyenne de la distribution. Il correspond à une mesure de l'incertitude qui est sans dimension. Les « boxplot » de la Figure 5.2 présentent la répartition du rapport de la médiane sur la valeur déterministe. Puis ceux de la Figure 5.3 présentent la répartition des coefficients de variation sur l'ensemble des processus de la base de données.

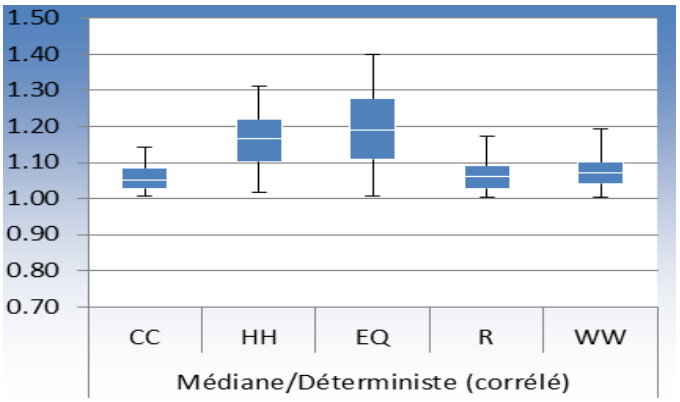


Figure 5.2 Boxplots des ratios des médianes sur les valeurs déterministes dans le cas corrélé.

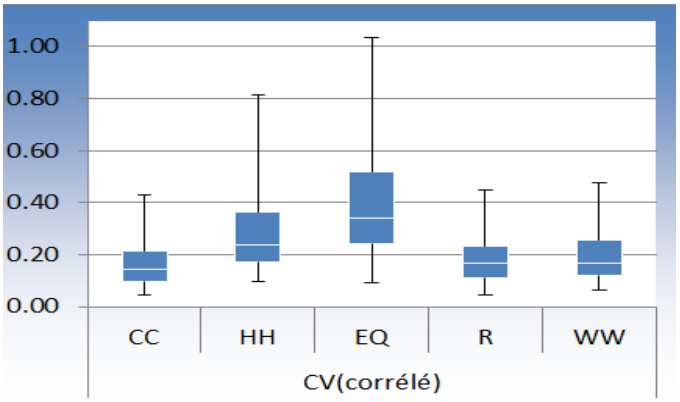


Figure 5.3 Boxplots des coefficients de variation dans le cas corrélé

Les boxplots de la Figure 5.2 sont construits de la manière suivante : la médiane de la distribution est représentée par la ligne blanche, la boîte est limitée par le 1er et le 3^{e} quartile et les barres sont limités par les 5^{e} et 95^{e} percentiles. Ces boxplot montrent que la médiane de la distribution est légèrement plus élevée que la valeur déterministe. Cela signifierait que le calcul déterministe sous-estime les impacts. Le graphe montre aussi que dans les catégories de dommages de la santé humaine et de la qualité des écosystèmes, ce phénomène est plus important, mais aussi moins systématique.

Les boxplots de la Figure 5.3 montrent que le coefficient de variation se place autour 10% et 50% pour la majorité des cas. Cependant, les résultats sont relativement étalés et 5% des processus ont un coefficient de variation supérieur à 100% pour la qualité des écosystèmes.

5.1.2 Propagation de l'incertitude dans le cas décorrélé

Dans un second temps, la propagation des incertitudes dans le cas partiellement non corrélé est étudiée. Le critère de désagrégation qui limite la progression de l'arbre tronqué est pris à 1% des impact (le maximum des catégories de dommages est considéré). Tout comme le cas précédent, une simulation de Monte-Carlo est faite sur les résultats d'impacts sur les 5 catégories de dommages. Les résultats obtenus sont relativement similaires. Tout d'abord, il a été vérifié que pour les deux méthodes de calcul (corrélé et non corrélé) les valeurs déterministes étaient parfaitement égales.

Les boxplots de la Figure 5.4 montrent que, statistiquement, le rapport entre la médiane de la distribution et la valeur déterministe supérieur à 1. Mais la disparité des valeurs est aussi plus grande dans les catégories de santé humaine et de qualité des écosystèmes et le rapport reste toujours légèrement supérieur à 1. Ensuite, les boxplots de la Figure 5.5 montrent que les coefficients de variation toujours entre 10% et 50% pour la moitié des processus.

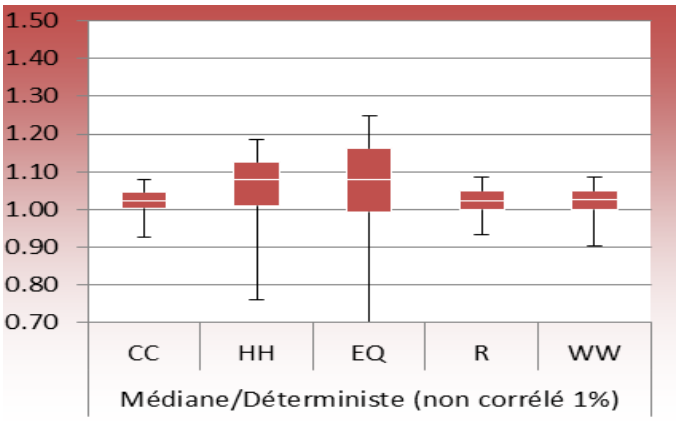


Figure 5.4 Boxplots des médianes sur les valeurs déterministes dans le cas non corrélé.

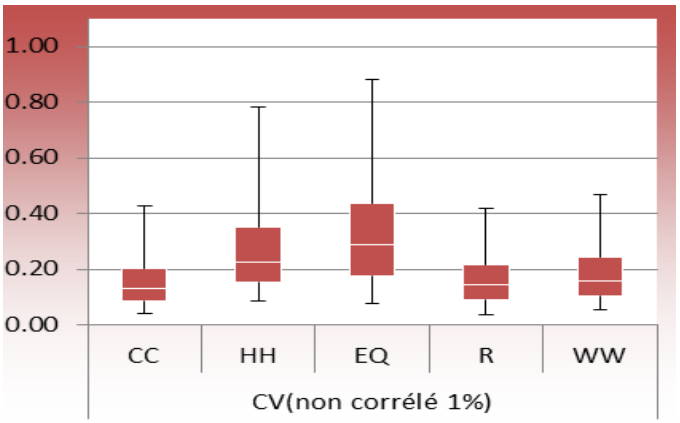


Figure 5.5 Boxplots des coefficients de variation dans le cas non corrélé.

Afin de s'assurer que le critère de désagrégation était suffisant, la part des impacts du système qui sont effectivement décorrélé a été observée. Plus précisément, c'est le pourcentage de la part des impacts des processus non agrégés dans l'impact des systèmes de produits complets qui est comparé au pourcentage de la part des impacts des processus agrégés. Les figures 5.6 et 5.7 représentent ces résultats dans des boxplots limités par les 1er et 3^{e} quartiles.

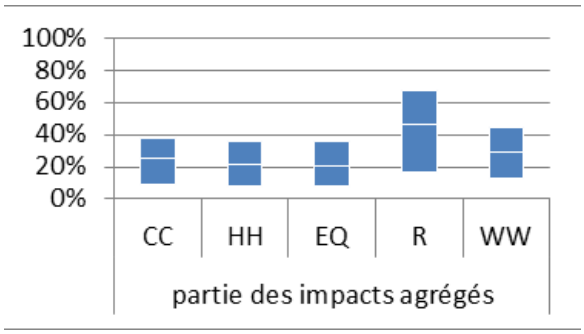


Figure 5.6 Boxplots des pourcentages des impacts agrégés dans les impacts totaux du système.

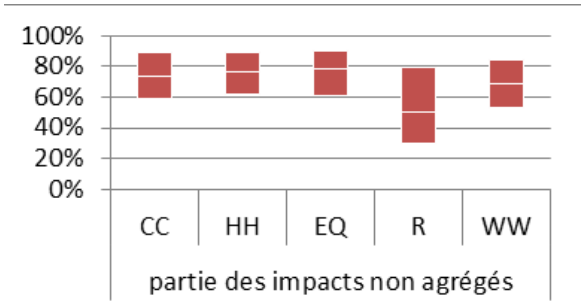


Figure 5.7 Pourcentage des impacts non agrégés dans les impacts totaux du système.

Ces graphes montrent bien que pour un critère de désagrégation de 1%, la part des impacts non agrégés dans le système (c'est-à-dire les impacts des processus non corrélés) représente la majorité des impacts du système total. On peut voir que pour toutes les catégories de dommage hormis les ressources, la médiane est aux alentours de 70% pour les impacts des processus non corrélés.

5.1.3 Comparaison des deux incertitudes

À présent, les cas corrélé et non corrélé sont comparés. Tout d'abord, le rapport de la première variable est observé. Cela correspond au rapport des médianes, car les valeurs déterministes sont égales dans les deux cas. Les boxplots de la Figure 5.8 présentent le rapport des médianes du cas corrélé sur celles du cas non corrélé. La médiane de ces rapports sur tous les processus est très proche de 1, mais toujours au-dessus. Cela semble montrer que le décalage entre la médiane et la valeur déterministe est moins important dans le cas non corrélé que dans le cas corrélé.

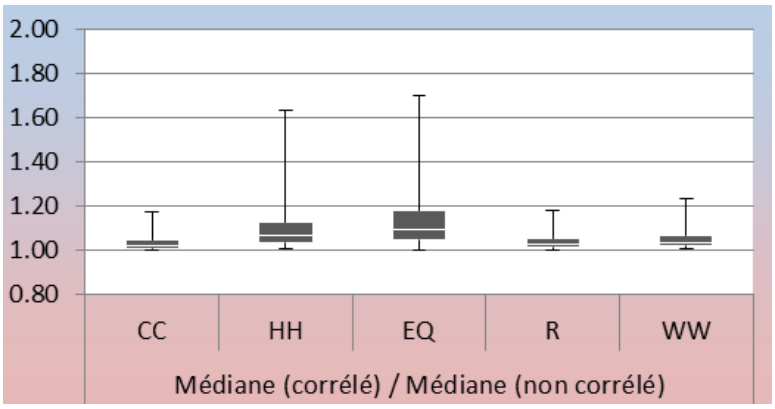


Figure 5.8 Boxplots des rapports des médianes du cas corrélé sur celles du cas non corrélé.

Ensuite, les Boxplots de la Figure 5.9 montrent bien que généralement le rapport du coefficient de variation de cas corrélé sur celui du cas non corrélé est supérieur à 1. Cela signifie que pour 75% des processus, les incertitudes des résultats d'impacts sont plus importantes dans le cas du calcul avec l'hypothèse de corrélation par rapport au cas sans cette hypothèse.

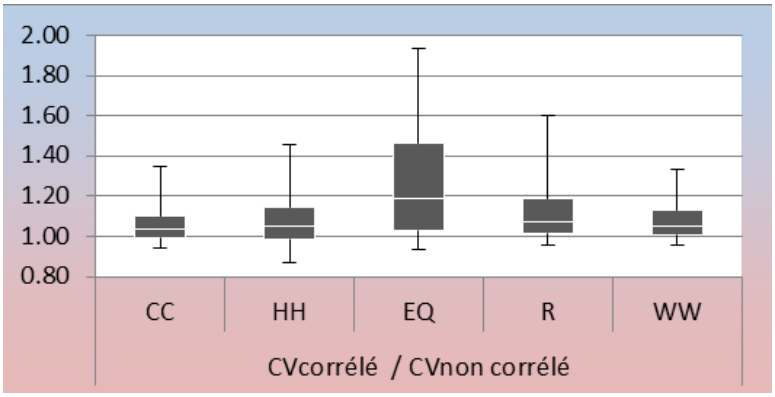


Figure 5.9 Boxplots du rapport des coefficients de variations du cas corrélé sur ceux du cas non corrélé.

Cependant, ces boxplot montrent aussi que les rapports des coefficients de variations sont relativement dispersés. Cela signifie que tous les processus n'ont pas la même sensibilité à l'hypothèse de corrélation. De plus, un nombre non négligeable de processus ont un rapport en dessous de 1, cela signifie que le résultat obtenu est contraire au résultat attendu dans ces cas. En effet, si le rapport est inférieur à 1, alors les impacts dans le cas du calcul corrélé seront moins incertains que ceux du cas non corrélé. Ce qui est contre-intuitif, car le fait de tirer des valeurs différentes pour chaque apparition d'un même processus devrait resserrer l'impact. En effet, les variables étant aléatoires, une valeur tirée un peu plus grande viendrait annuler une tirée un peu plus petite.

La Figure 5.10 présente la comparaison des résultats des simulations de Monte-Carlo pour un système de produit arbitraire. Le système de produit présenté est la production de 1 kWh de « *Electricity, hydropower, at power plant/FR U* ».

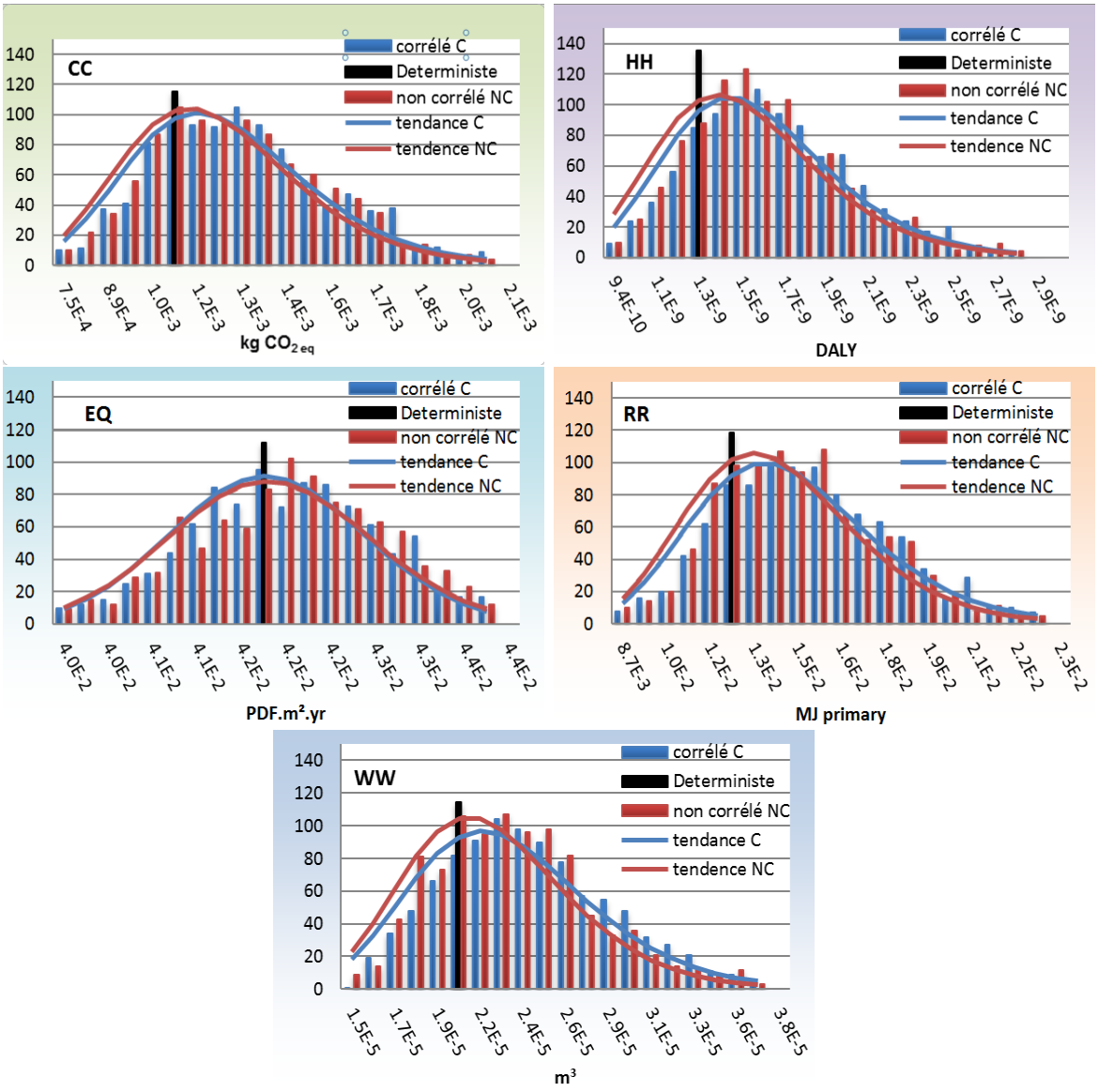


Figure 5.10 Histogrammes et tendances dans les cas corrélé et non corrélé pour la production de 1 kWh de « *Electricity, hard coal, at power plant/UCTE U* ».

Dans ces graphes, les deux échantillonnages sont très similaires. De plus, l'approximation Log-normale est fidèle à la distribution empirique. Dans ce cas, la valeur déterministe apparaît comme légèrement plus faible que le mode de la distribution, ce qui vient confirmer les résultats précédents. Cependant, ce phénomène est observé aussi bien dans le cas corrélé que dans le cas non corrélé. Finalement, il est difficile de distinguer dans ce cas quel est le cas

le plus incertain. Les distributions Log-normales approchées sembenlt avoir des paramètres d'incertitudes relativement proches.

5.1.4 Étude des causes communes de sensibilités

À présent, les causes de la sensibilité de l'hypothèse de corrélation intradonnées sont étudiées. Tout d'abord, les catégories de produits pour lesquels l'hypothèse de corrélation intradonnées est la plus importante (au-dessus de 150% ou au-dessous de 30%) sont listées. Le Tableau 5.1 donne les pourcentages des produits dont la sensibilité dépasse les bornes précédentes pour certaines catégories de processus.

Tableau 5.1 Catégories de produits les plus sensibles à l'hypothèse de corrélation intrasystème.

	Categories de produit	**% de produit en dehors de [0.3 ; 1.50]**	**nbr de processus en dehors de [0.3 ; 1.50]**
CC	material / Chemicals / Washing agents	17%	5
	material / Chemicals / Pesticides	35%	33
	material / Chemicals / Acids (inorganic)	11%	2
	material / Construction / Ventilation	10%	4
	material / Glass / Packaging	23%	3
	material / Textiles	14%	1
HH	energy / Electricity by fuel / Nuclear	24%	6
	material / Chemicals / Pesticides	11%	10
	material / Chemicals / Washing agents	21%	6
	material / Construction / Cladding	100%	1
	material / Construction / Doors	50%	2
	material / Electronics / Photovoltaic	27%	8
	material / Fuels / Uranium	10%	6
	material / Glass / Construction	33%	4
	material / Glass / Packaging	23%	3
	material / Metals / Extraction	17%	1
	material / Textiles	29%	2
	processing / Glass	50%	1
EQ	energy / Electricity by fuel / Nuclear	16%	4
	energy / Electricity by fuel / Photovoltaic	42%	27
	energy / Electricity country mix / Production	16%	6
	material / Chemicals / Washing agents	10%	3
	material / Glass / Packaging	38%	5
R	material / Agricultural / Plant oils	13%	2
	material / Agricultural / Plant production	30%	35
	material / Chemicals / Pesticides	13%	12
	material / Chemicals / Washing agents	10%	3
	material / Electronics / Photovoltaic	13%	4
	material / Textiles	14%	1
	processing / Textiles	20%	1
WW	energy / Cogeneration / Oil	42%	5
	energy / Mechanical / Infrastructure	100%	1
	material / Agricultural / Plant production	15%	17
	material / Chemicals / Washing agents	31%	9
	material / Construction / Coverings	12%	2
	material / Construction / Insulation	10%	2
	material / Construction / Ventilation	15%	6
	material / Glass / Packaging	31%	4
	material / Metals / Ferro	39%	7
	processing / Agricultural / Operations	32%	7
	waste treatment / Landfill / Sanitary landfill	10%	3

Il est possible de retirer de ce tableau que la sensibilité à l'hypothèse de corrélation intradonnées se retrouve beaucoup dans des processus associés à la catégorie principale « material » et donc des matériaux. Plus précisément, les matériaux chimiques sont relativement présents avec notamment les agents nettoyants ou encore les pesticides. Dans la catégorie de dommage de la qualité de l'écosystème, on retrouve aussi des processus associés à l'énergie électrique provenant de l'essence.

Finalement, il est intéressant de connaître la corrélation entre la sensibilité de l'hypothèse de corrélation intradonnées et la désagrégation des arbres. Comme présenté dans la méthodologie, un score de désagrégation est calculé pour chacun des processus de la base de données dans chacun des arbres représentant un système de produit dans le cas non corrélé. Le score est le produit du nombre de réapparitions d'un processus dans l'arbre et de sa contribution relative eux impacts totaux par unité de produit. Ensuite, la corrélation entre le score de désagrégation pour le processus et la sensibilité est calculée. Cela permet de connaître les processus dé-corrélés qui sont les plus responsables de la sensibilité à l'hypothèse.

Le Tableau 5.2 présente les processus qui ont les plus grandes corrélations (supérieurs à 8%) entre leurs scores de désagrégations à travers les arbres et la sensibilité à l'hypothèse de corrélation intradonnées dans ces arbres. Il est tout de même observable que les corrélations sont relativement faibles. Cela signifie que le score de désagrégation n'est pas une explication suffisante aux variations de la sensibilité selon les arbres.

Tableau 5.2 Processus avec la plus grande corrélation du score de désagrégation et de la sensibilité de l'hypothèse.

	Processus	$\rho_{\mathbf{s},\alpha}$
CC	Disposal, building, cement (in concrete) and mortar, to sorting plant/CH U	28%
	Disposal, building, glass pane (in burnable frame), to final disposal/CH U	28%
	Disposal, building, reinforcement steel, to sorting plant/CH U	26%
	Sulphur hexafluoride, liquid, at plant/RER U	15%
	Charcoal, at plant/GLO U	15%
HH	Natural gas, sour, burned in production flare/m3/GLO U	12%
	Water, decarbonised, at plant/RER U	11%
	Electricity, production mix IT/IT U	11%
	Light emitting diode, LED, at plant/GLO U	11%
	Diode, glass-, SMD type, surface mounting, at plant/GLO U	11%
EQ	Natural gas, sour, burned in production flare/m3/GLO U	11%
	Hard coal, at regional storage/ZA U	10%
	Hard coal, at regional storage/EEU U	9%
	Hard coal, at regional storage/RLA U	9%
	Hard coal, at regional storage/WEU U	9%
R	Natural gas, sour, burned in production flare/m3/GLO U	13%
	Electricity, production mix CZ/CZ U	11%
	Electricity, production mix GR/GR U	11%
	Electricity, production mix IT/IT U	11%
	Electricity, production mix PL/PL U	11%
WW	Electricity, production mix IT/IT U	14%
	Electricity, production mix PL/PL U	13%
	Electricity, production mix ES/ES U	13%
	Electricity, production mix GR/GR U	13%
	Electricity, production mix CZ/CZ U	13%

5.2 Corrélation intersystème dans l'incertitude d'un système de produits agrégés

Dans cette partie, c'est la corrélation au niveau intrasystème qui est étudiée. C'est-à-dire la corrélation entre les différents produits intervenant dans un système de produits constitués de produits agrégés. Les incertitudes de chacun des systèmes de processus sont connues, celles-ci ont été déterminées par les simulations de Monte-Carlo de la partie précédente. Il est alors intéressant de remarques que les impacts peuvent être choisis avec des incertitudes qui prennent en compte la corrélation intradonnées, ou qui ne la prennent pas en compte.

Cependant pour connaître les incertitudes des résultats d'impact d'un système de produit composé de systèmes de produits agrégés, il faut non seulement connaître les incertitudes de

chacun, mais aussi les covariances entre ces différents systèmes de produits agrégés.

5.2.1 Matrices de variances-covariances de tous les processus unitaires.

Les matrices de variances-covariances sont alors estimées grâce aux échantillons des simulations de Monte-Carlo de la section 5.1.1 (cas corrélé) comme présenté dans la méthodologie. Les graphiques de la Figure 5.11 montrent la répartition des valeurs de corrélation dans chacune des catégories de dommage à travers des histogrammes. Ces derniers nous permettent de remarquer que les corrélations ne sont absolument pas nulles et donc que les covariances entre les impacts agrégés des processus élémentaires ne sont pas négligeables devant leurs variances. Cela nous permet de savoir que ces impacts ne sont pas indépendants entre eux.

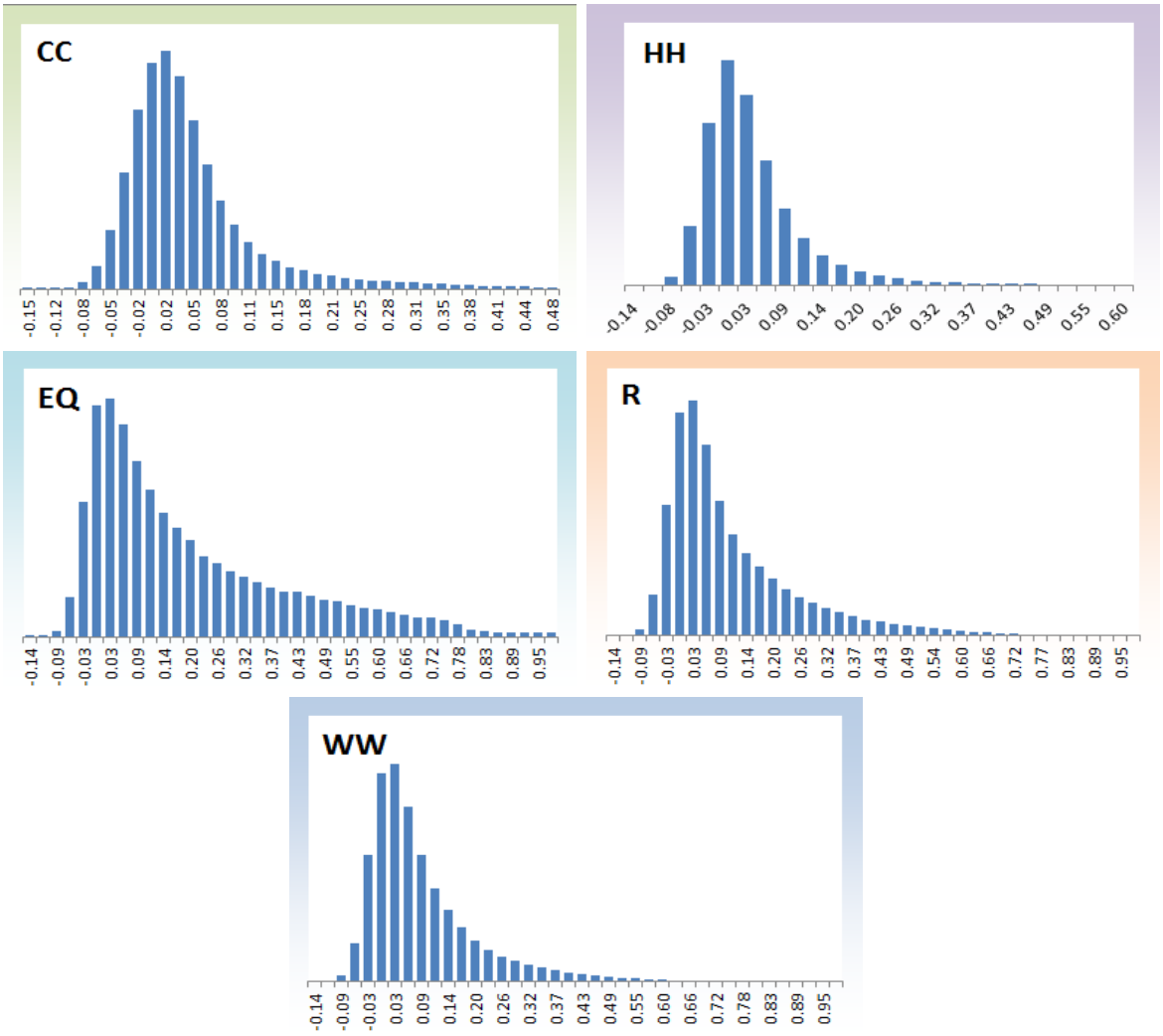


Figure 5.11 Distributions empiriques des coefficients de corrélations entre les différents systèmes de produits agrégés.

Les histogrammes révèlent qu'une partie non négligeable des corrélations sont négatives. Cela signifie que certains impacts agrégés varient de manières opposées d'un processus élémentaire à l'autre. Cela peut par exemple être provoqué par les énergies. Car le fait d'utiliser une certaine source d'énergie réduit l'utilisation d'une autre. Donc si les impacts associés aux différentes sources sont significativement différents, alors les variations seront opposées et les impacts des processus seront corrélés négativement. Cependant, comme les histogrammes le montrent, cela arrive plutôt rarement.

5.2.2 Calcul de l'incertitude d'un système de produits agrégés non indépendants.

Les données d'incertitudes sur les produits agrégés ainsi que les matrices de variances-covariances permettent alors de déterminer la variance d'un système de produits composé de produits agrégés. Pour cette étude, ce sera le système correspondant à une étude de cas proposée par le partenaire industriel Nestlé qui est utilisé. Pour des raisons de confidentialités, le détail du système n'est pas présenté ici, seuls les résultats apparaissent dans le Tableau 5.3. Dans ce tableau, les résultats en considérant les covariances et sans considérer les covariances (les produits agrégés sont considérés indépendants) sont comparés.

Tableau 5.3 Résultats d'impacts avec incertitude pour un système de produits agrégés (système 1 provenant de l'étude de cas).

	Valeur déterministe	Mode	Écart type (corrélés)	Écart type (indépendants)
CC (10^{-2}kg$_{CO_2eq}$)	14.5	14.8	1.29	1.15
HH (10^{-8} DALY)	9.17	10.3	0.953	0.814
EQ (10^{-2} PDF.m^2.yr)	17.1	17.5	0.901	0.746
R (MJ primary)	2.86	2.87	0.301	0.249
WW (10^{-2} m^3)	11.7	11.7	0.234	0.231

5.2.3 Part des covariances dans l'incertitude du système total.

Il est possible de connaître la part de la covariance dans le calcul de la variance en comparant la variance dans les deux cas. En général, la part de la covariance dans le calcul de la variance totale n'est pas négligeable. Cette proportion dépend du système et de la corrélation des différents produits intervenants dans le système. Le Tableau 5.4 présente le pourcentage de la part de la covariance dans la variance totale associée aux impacts de l'étude de cas (impacts du cycle de vie d'un burger ...). Il est clair que dans la majorité des catégories de dommages, ce pourcentage n'est pas négligeable.

Tableau 5.4 Part de la covariance dans la variance d'un système de produits agrégés (système 1 provenant du l'étude de cas).

	Variance totale	**Covariances**	**Covariances/Variance**
CC (10^{-5} $kg^2_{CO_2eq}$)	16.7	3.42	20 %
HH (10^{-17} $DALY^2$)	9.09	2.45	27 %
EQ (10^{-5} $PDF^2.m^4.yr^2$)	8.11	2.55	31 %
R (10^{-2} $MJ^2_{primary}$)	9.06	2.88	32 %
WW ($10^{-7}m^6$)	54.8	1.33	2 %

5.3 Corrélation intersystème dans la comparaison de systèmes de produits agrégés.

La corrélation entre les différents produits agrégés qui composent un même système est à prendre en compte aussi lors de la comparaison de plusieurs produits. Par exemple, une entreprise produisant deux produits (sortants) va probablement utiliser les mêmes produits (entrants) pour les deux fabrications : même camion pour transporter les deux produits, même film plastique pour les emballages, mêmes sources de matières premières ou d'énergies...

Il est alors important de considérer ces corrélations lors de la comparaison des impacts.

5.3.1 Comparaison de deux systèmes en considérant leurs corrélations

Tout comme dans la partie précédente, l'étude se base sur deux systèmes de produits qui sont fournis par le partenaire industriel et présentés dans le tableax 4.1.

Comme il a été présenté dans le chapitre de méthodologie, un système est construit avec la différence entre les produits du système 1 et ceux du système 2. Ensuite, une hypothèse de normalité est faite sur les distributions des impacts de ce nouveau système. Enfin, les paramètres de la distribution sont estimés grâce au calcul de la moyenne et de la variance. Ces deux paramètres sont estimés de la manière suivante : la moyenne correspond à la somme des moyennes des lois Log-normales qui composent le système, pondérées des quantités de chaque produit. Ensuite, l'écart type est obtenu comme racine de la variance qui elle-même est obtenue comme la variance de la somme des impacts qui composent le système de la différence (cette variance prend en compte les covariances entre tous les différents produits qui composent les deux systèmes à comparer).

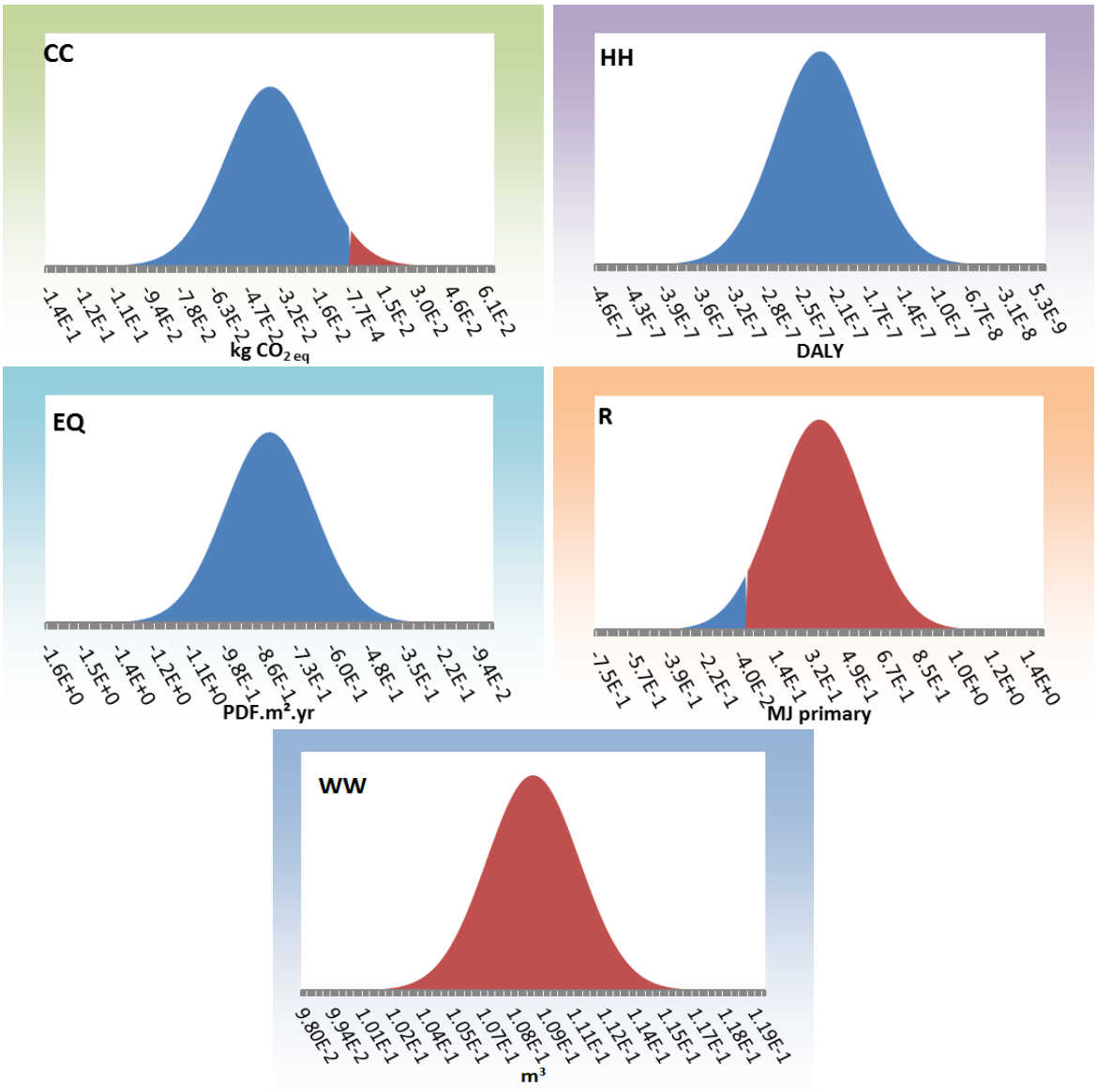


Figure 5.12 Distributions de la différence de deux systèmes autour de 0.

La Figure 5.12 présente les distributions normales de l'impact du système correspondant à la différence des deux systèmes comparés. Ainsi les probabilités que l'impact soit inférieur à 0 correspond à la probabilité que le système 1 soit moins impactant que le système 2, et vice et versa. Dans cette étude de cas, il est donc difficile de préférer un système à l'autre, car le résultat dépend des catégories de dommage. Par exemple, en santé humaine et en qualité des écosystèmes, le système 1 (bleu) sera certainement moins impactant. En revanche en prélèvement des eaux, c'est le contraire. De plus, dans les catégories de ressources et changements climatiques, le résultat n'est pas certain, mais il est fort probable avec des conclusions inverses.

Il est de plus intéressant de savoir quelle est la corrélation entre le système 1 et le système 2. Pour cela, les formules suivantes sont utilisées (X_1 et X_2 représentent respectivement les variables aléatoires associées aux impacts des systèmes 1 et 2) :

$$Cov(X_1, X_2) = \frac{Var(X_1) + Var(X_2) - Var(X_1 - X_2)}{2}$$

$$\rho_{X_1,X_2} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}}$$

Tableau 5.5 Corrélations entre les systèmes comparés de l'étude de cas (X_1 : système 1, X_2 : système 2).

	ρ_{X_1,X_2}
CC (Kg CO_2 eq)	23%
HH (DALY)	24%
EQ (PDF.m^2.yr)	55%
R (MJ primary)	68%
WW (m^3)	56%

Le Tableau 5.5 montre bien que les impacts des deux systèmes de l'étude de cas sont relativement corrélés. Ces résultats s'expliquent par le fait que certains produits sont utilisés dans les deux systèmes et sont donc totalement corrélés. Ces mêmes produits correspondent par exemple au transport qui amènerait les deux systèmes de produits (burgers) au distributeur, les impacts associés à ce transport sont donc totalement corrélés pour les deux systèmes.

5.3.2 Permettre à l'utilisateur de spécifier l'indépendance des mêmes produits utilisés.

Il est intéressant de permettre à l'utilisateur du logiciel de choisir si un même produit utilisé dans les deux systèmes est corrélé ou non. Par exemple, lors de la comparaison de deux recettes de gâteaux, si l'ingrédient de sucre pour les deux recettes provient du même fournisseur, alors il est corrélé lors de l'analyse d'incertitude. Cependant, si l'ingrédient provient de deux sources différentes pour chacun des gâteaux, alors il ne doit pas être corrélé. Dans la situation actuelle, les deux apparitions du produit seront automatiquement corrélées. C'est pourquoi on permet à l'utilisateur, lors de la construction du système, de choisir si les différentes apparitions d'un même produit sont corrélées ou non.

5.4 Intégration de l'étude au logiciel d'écodesign Ecodex.

Un des buts de l'étude est de pouvoir utiliser les résultats obtenus dans les logiciels d'écodesign des industriels, et plus précisément dans celui du partenaire industriel. Afin de faciliter une intégration future des résultats dans les logiciels, un outil développé sous le langage de programmation **Python** et intègre les résultats de l'étude. Cet outil sert de preuve de concept et facilitera ensuite l'intégré au logiciel du partenaire par ses soins. Le code de l'outil peut aisément être transcrit dans n'importe quel langage de programmation comme **PHP**, **Java** ou **C++** par exemple.

5.4.1 Fournir une information pertinente au meilleur choix de système.

Il est important pour les industriels d'accéder aux résultats de manière simple. C'est pourquoi la transformation de la section précédente qui permet de passer de la distribution à un simple graphe en barres est appliquée. De plus, les impacts des deux systèmes sont présentés sous un autre graphe permettant de les analyser de manière indépendante et séparés. Le graphe de la Figure 5.13 permet de connaître les impacts relatifs des deux systèmes, mais aussi de connaître les incertitudes relatives aux impacts dans chacune des catégories de dommages.

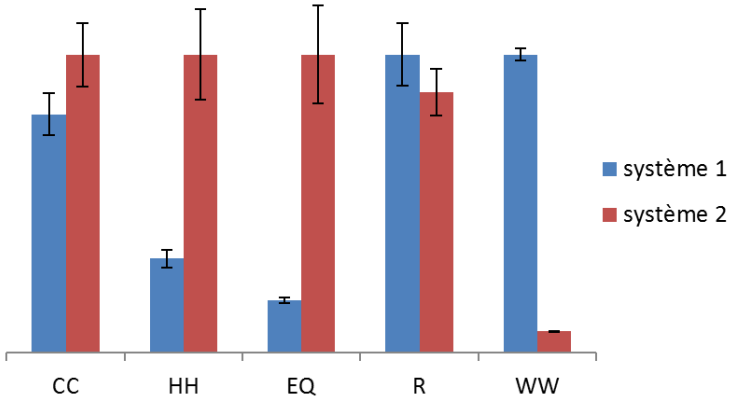


Figure 5.13 Graphe de comparaison des impacts de deux systèmes.

Ce graphe est accompagné d'un tableau comprenant pour chacune des catégories de dommages la valeur déterministe du calcul de l'impact, le mode de la distribution des impacts, l'écart type et le coefficient de variation. Ainsi, grâce à une brève analyse des graphes de la Figure 5.13 et de la Figure 5.12, l'utilisateur peut comparer les impacts de deux systèmes de produits. De plus, s'il a besoin d'informations plus précises relatives à un système en particulier, l'information nécessaire se trouve dans les tableaux.

De plus, lors de la comparaison de deux systèmes de produits agrégés, ce ne sera pas les distributions qui seront présentées, mais une barre divisée en deux parties : une correspondant à la probabilité que la différence des impacts des deux systèmes de produits agrégés soit inférieure à 0 et l'autre à la probabilité qu'elle soit supérieure à 0. Autrement dit, une partie correspond à la probabilité que le système 1 soit moins impactant que le système 2 et l'autre partie la probabilité de la situation contraire. La Figure 5.14 donne une représentation de ces barres.

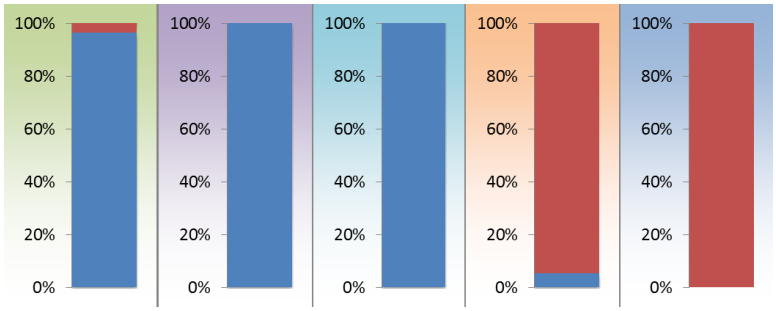


Figure 5.14 Graphe d'aide à la décision présent dans l'outil.

5.4.2 Développement d'un outil et application à une étude de cas.

L'outil est développé sous le langage **Python**, car c'est le langage de programmation qui a été utilisé jusqu'à présent durant l'étude. Cet outil fait appel à différentes bases de données qui sont simplement des fichiers « .csv ». Plus précisément, l'outil a besoin pour fonctionner de l'ensemble des noms et unités des produits de la base de données **ecoinvent v2.2**. Ensuite, les impacts sont répertoriés pour chaque produit agrégé de la base de données. La valeur déterministe des impacts et les paramètres de la distribution sont stockés dans chacune des catégories de dommages et pour les deux cas de corrélation intradonnées (totalement corrélé et non corrélé). Cela permet à l'outil de différencier les deux cas. Ensuite, il sera possible de préparer deux systèmes de produits qui seront inscrits dans deux fichiers « system1.csv » et « system2.csv ». Cela permet à l'utilisateur d'importer par exemple deux systèmes sans avoir à les construire entièrement dans l'outil. Finalement, les matrices de variances pour chaque catégorie de dommage sont aussi stockées. Cependant, ce ne sont pas des matrices, car toutes les valeurs inutiles ont été enlevées (les diagonales et toutes les valeurs au-dessus de la diagonale). Des listes de listes de longueurs différentes composent alors ces fichiers. Cela permet un gain de mémoire de stockage non négligeable et une rapidité d'ouverture accrue.

L'outil permet donc soit d'importer directement deux systèmes pré-construits, soit de construire deux systèmes de produits en utilisant la base de données des produits agrégés **ecoinvent**

v2.2. Il est aussi possible de choisir d'utiliser les données d'incertitudes issues des résultats sans l'hypothèse de corrélation intradonnées, ou avec l'hypothèse de corrélation intradonnées. L'image de la Figure 5.15 présente l'interface de l'outil et ses différentes fonctionnalités. Un système de recherche intelligente est mis en place pour rechercher les processus dans la base de données. Il suffit d'indiquer une série de mots clés apparaissant dans le nom du processus, et une liste restreinte des processus correspondants est proposée. Ainsi, les systèmes à étudier et à comparer peuvent être construits. La base de données peut être choisie parmi les résultats du cas corrélé (corrélation intradonnées) et le cas non corrélé (corrélation intradonnées). Puis le calcul peut être lancé.

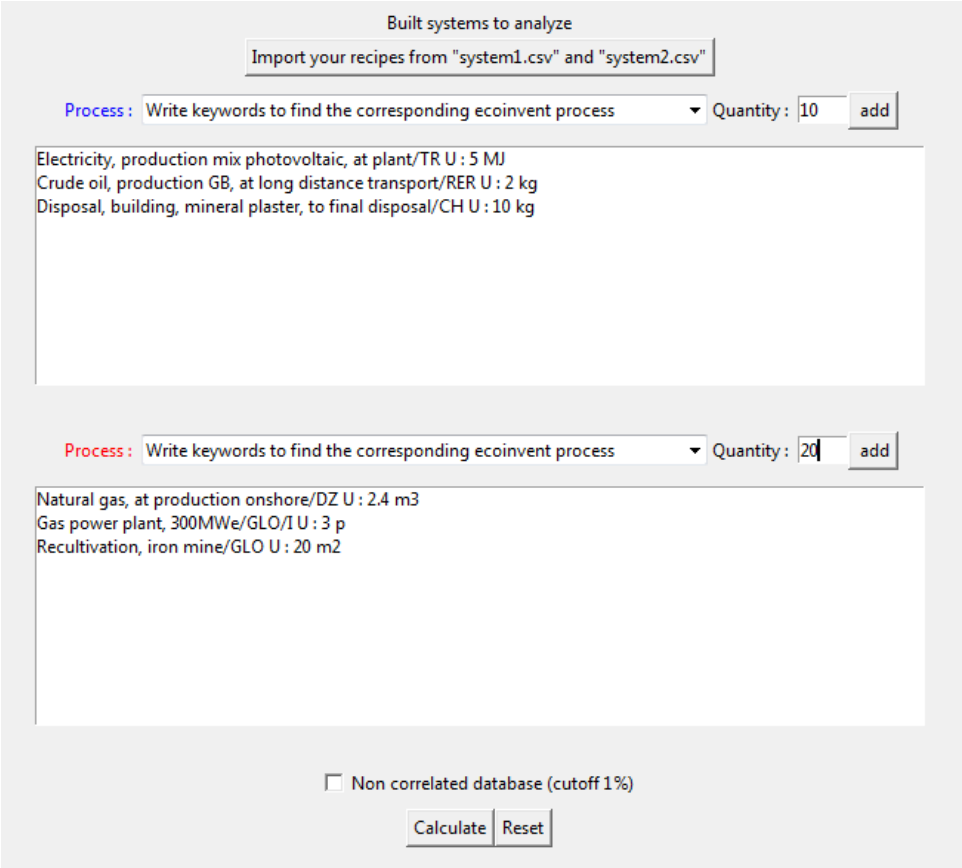
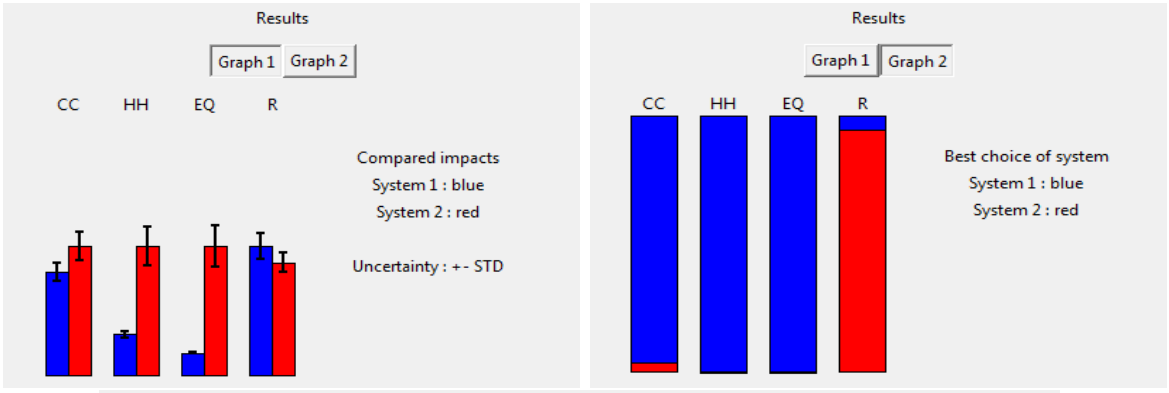


Figure 5.15 Interface de construction des systèmes dans l'outil développé.

Calcul des impacts AVEC incertitudes et considération de plusieurs niveaux de corrélations.

L'outil prend en compte les différents niveaux de corrélation. Tout d'abord dans le choix de la base de données pour la corrélation intradonnées. Puis lors du calcul des résultats des

impacts. Les résultats des études précédentes sont intégrés dans l'outil. Le calcul des incertitudes (calcul de la variance) utilise les fichiers de covariances pour intégrer les corrélations intrasystème et intersystème aux résultats. Ensuite les résultats exposés dans l'outil sont ceux présentés précédemment. Ils sont présentés dans la Figure 5.16.



	Determinist	Mode	SD	CV
CC (kg CO2 eq)	1.45e-01	1.48e-01	1.29e-02	8.75e-02
HH (DALY)	9.17e-08	1.03e-07	9.53e-09	9.28e-02
EQ (PDF*m2*yr)	1.71e-01	1.75e-01	9.01e-03	5.15e-02
R (MJ primary)	2.863	2.871	3.01e-01	1.05e-01
WW (m3)	1.17e-01	1.17e-01	2.34e-03	2.00e-02

	Determinist	Mode	SD	CV
CC (kg CO2 eq)	1.84e-01	1.85e-01	1.96e-02	1.06e-01
HH (DALY)	3.07e-07	3.26e-07	4.94e-08	1.52e-01
EQ (PDF*m2*yr)	1.02	1.0	1.64e-01	1.64e-01
R (MJ primary)	2.465	2.511	2.26e-01	8.99e-02
WW (m3)	8.25e-03	8.34e-03	2.49e-04	2.98e-02

Figure 5.16 Présentations des résultats dans l'outil.

Réponds aux exigences de rapidités de calcul

L'outil permet alors de répondre aux attentes de l'étude. De plus, certaines contraintes doivent être respectées comme la rapidité d'exécution des calculs. L'outil nous permet de vérifier que le calcul est extrêmement rapide : une fraction de seconde. À l'ouverture de l'outil, les bases de données sont chargées et cela prend quelques secondes (peut aller jusqu'à une minute

selon la performance de la machine). Cependant, l'utilisation de calcul parallèle permet de charger les bases de données pendant que l'utilisateur construit le système. Ainsi, le temps de chargement n'est pas perceptible par l'utilisateur.

Facile à intégrer dans le logiciel d'écodesign du client.

Le développement de cet outil en python permet d'avoir une application concrète qui nous prouve que les résultats de l'étude sont intégrables à un programme. Aucune librairie spécifique au langage de programmation n'est utilisée hormis « **tkinter** » qui permet d'avoir une interface graphique et qui a des équivalents dans la plupart des langages de programmation. L'outil est alors fourni au partenaire pour soit en faire une utilisation directe, pour y analyser le code qui est commenté et explique afin de l'intégrer dans leur logiciel existant d'écodesign **Ecodex**.

CHAPITRE 6 DISCUSSION

Dans ce chapitre, les apports de l'étude au domaine de l'ACV ainsi que ses limites sont discutés. En réponse au bilan de la revue de la littérature de la section 2.4, il est vu comment l'étude réalisée complète la recherche déjà effectuée en propagation de l'incertitude en analyse de cycle de vie.

6.1 Les apportes de l'étude au domaine de recherche.

Tout d'abord, la sensibilité à l'hypothèse de corrélation intradonnées n'avait pas encore été explorée. L'étude apporte donc une piste de solution qui permettrait de s'approcher de la réalité en ne faisant pas cette hypothèse. En effet, l'algorithme de désagrégation permet de décorréler certains processus et les analyses de Monte-Carlo nous permettent ensuite de comparer la propagation des incertitudes dans les cas corrélé et non corrélé. Finalement cette hypothèse de corrélation intradonnées n'introduit pas une erreur très importante. L'étude révèle que la sensibilité de l'hypothèse est relativement faible pour une grande partie des systèmes de processus élémentaires.

Ensuite, les simulations de Monte-Carlo réalisées pour l'étude donnent des informations sur les incertitudes des systèmes de produits qui peuvent donc compléter les bases de données agrégées. Une idée des distributions des impacts au niveau des dommages pour chacun des processus élémentaires de la base de données **ecoinvent v2.2** est obtenue. Cela donne donc accès à une approximation de la loi qui semble généralement correspondre à une loi Log-normale et parfois à une loi Normale.

Finalement, ces mêmes simulations de Monte-Carlo ont permis de déterminer une approximation des corrélations entre les différents systèmes de produits. Ces résultats peuvent améliorer les analyses d'incertitudes lors de la comparaison de systèmes de processus ou encore déterminer plus précisément les incertitudes associées aux impacts d'un système de produits agrégés.

6.2 Les limites des apports de l'étude.

6.2.1 La sensibilité de l'hypothèse de corrélation intradonnées.

En ce qui concerne l'hypothèse de corrélation intradonnée, la sensibilité de cette hypothèse a été analysée et une origine de cette sensibilité a tenté d'être déterminée. Cependant, la réalité

se trouve entre le cas totalement corrélé et le cas totalement décorrélé. C'est pourquoi une amélioration du modèle utilisé pourrait être apportée. Par exemple améliorer l'algorithme de désagrégation qui consisterait à stopper l'arborescence de l'arbre non seulement lorsque l'impact est négligeable, mais aussi à l'apparition d'un produit qui est supposé toujours corrélé. Ou encore en arrêtant la décorrélation avec un critère sur les incertitudes en non les impacts.

Il est aussi possible d'améliorer cet algorithme en ne faisant pas une hypothèse de log-normalité pour les impacts associés aux données agrégées des nœuds finaux. Il serait possible d'utiliser différentes lois (log-normale, normale, ou triangulaires) pour les impacts au niveau dommage selon le processus élémentaire.

Enfin, il pourrait y avoir d'autres sources d'explications à la sensibilité de l'hypothèse de corrélation intradonnées. Par exemple, le score de désagrégation qui est défini dans la section 4.1.4 peut être amélioré pour prendre en compte d'autres caractéristiques de la désagrégation qui expliquerait mieux la sensibilité de cette hypothèse.

6.2.2 L'incertitude des impacts de systèmes de produits agrégés.

En ce qui concerne la prise en compte des données d'incertitudes dans l'utilisation de systèmes de produits agrégés. L'étude ne fournit que des résultats sur l'incertitude globale. En effet, seulement la variance est calculée pour donner les informations sous forme de barres d'erreurs. Il pourrait être intéressant de fournir des informations plus complètes comme un histogramme, ou une approximation de la distribution réelle par une distribution connue. Pour cela, il serait possible de stocker non pas les paramètres de la loi qui estime la vraie distribution, mais l'intégralité de l'échantillon obtenu à la suite de la simulation de Monte-Carlo pour chacun des systèmes de processus élémentaire. Ensuite, ces échantillons seraient directement utilisés afin de connaître les impacts d'un système de produits agrégés complet. Cette méthode serait sûrement plus longue et demanderait plus d'espace pour stocker les données, mais en contrepartie elle donnerait davantage d'information sur les incertitudes des résultats.

De plus, si la simulation de Monte-Carlo qui nous donne les échantillons des impacts au niveau dommage des processus élémentaires est réalisée comme dans la section 4.1.1 (cas totalement corrélé du réseau), c'est-à-dire en utilisant la matrice identité comme vecteur de demande, alors les informations de corrélations entre les impacts des différents processus élémentaires seront prises en compte dans ces échantillons. Cela permettrait, lors du calcul de l'incertitude d'un système de produits agrégé de connaître l'incertitude des impacts du système tout en ayant considéré les corrélations entre les différents produits agrégés qui le

composent.

Cette même solution permettrait de ne pas avoir à faire l'hypothèse de normalité pour considérer la corrélation intersystème lors de la comparaison de deux systèmes de produits. En effet, aucune hypothèse ne serait nécessaire. Une fois les impacts des échantillons combinés pour connaître l'échantillon correspondant aux impacts de la différence des deux systèmes de produit agrégé complets, il faut ensuite compter le pourcentage de résultats positifs dans l'échantillon pour savoir à quelle proportion un système est préférable à l'autre.

Cependant, cette solution ne pourrait pas être appliquée dans le cas des systèmes décorrélés, car chaque système possède sa propre matrice technologique et sa propre matrice intervention. Cela implique que les informations de corrélations intrasystème ne pourront pas être récupérées comme avec l'utilisation de la matrice identité comme vecteur de demande dans le cas totalement corrélé (corrélation intradonnées).

6.2.3 Nécessité de calculer les données d'incertitudes auparavant.

Dans la solution qui est proposée pour donner directement les informations d'incertitudes dans les logiciels ACV simplifiés, les informations d'incertitudes associées aux systèmes de processus élémentaires sont nécessaires. En effet, des simulations de Monte-Carlo sont réalisées en amont comme il l'a été expliqué dans la section 4.1. Cela implique que lorsque la base de données change, ces analyses doivent être entièrement refaites. De même si la méthode d'impact est modifiée. Les simulations de Monte-Carlo sont réalisées avec 10000 itérations, et donc un temps de calcul relativement long selon la taille de la base de données utilisée. De plus, l'utilisation de plusieurs bases de données ou de plusieurs méthodes d'impacts dans un même logiciel nécessiterait une grande quantité d'espace de stockage et une ouverture plus longue du logiciel, car chaque matrice de variance-covariance nécessiterait d'être stockée et ouverte au démarrage.

6.2.4 Des niveaux de corrélation non considérés.

Dans l'étude réalisée, trois niveaux de corrélations sont pris en compte lors de la propagation des incertitudes. Cependant, il existe davantage d'incertitudes et de corrélations qui n'ont pas été considérées ici. Par exemple les incertitudes des éléments de la matrice des facteurs de caractérisations. En effet, la matrice **CF** qui a été utilisée est considérée déterministe, cependant en réalité, les éléments qui la composent peuvent être aléatoires. Par exemple la méthode d'impact **IMPACT world +** considère ces incertitudes qui modifieraient l'incertitude des impacts d'un système de produit au niveau dommage.

De plus, certaines corrélations n'ont pas été considérées, comme la corrélation qui lie les différents entrants et sortants d'un processus. En effet, lors d'un processus de combustion de diesel, il y aura de la formation de CO_2. Les variations de diesel et de CO_2 sont corrélées dans la réalité alors que dans les simulations de Monte-Carlo, ces deux flux sont considérés comme indépendants.

CHAPITRE 7 CONCLUSION

Cette conclusion présente dans un premier temps une synthèse des travaux réalisés ainsi que leurs résultats. Puis dans un second temps, des recommandations sont évoquées.

7.1 Synthèse des travaux.

Les travaux réalisés dans cette étude visent à analyser la propagation des incertitudes dans les analyses de cycle de vie. Dans cette analyse, plusieurs niveaux de corrélations liés à cette propagation sont abordés. Il est tout d'abord abordé la corrélation intradonnées présente entre les mêmes produits qui apparaissent à différents endroits du cycle de vie d'un système de produit. Ensuite, la corrélation intrasystème qui correspond à la corrélation entre les impacts agrégés apparaissant dans un système de produit composé de produits agrégés est étudiée. Enfin, la corrélation intersystème est observée. Ce dernier niveau de corrélation correspond à la corrélation entre les impacts de deux systèmes de produits qui sont comparés.

La première partie de l'étude porte donc sur l'analyse de la sensibilité de l'hypothèse de corrélation intradonnées qui consiste à considérer comme corrélés les mêmes processus apparaissant à différentes positions de l'arbre du cycle de vie du produit. Afin d'étudier cette sensibilité, deux situations sont comparées. La première correspond à la situation de corrélation totale, dans laquelle l'arbre est représenté sous la forme d'un réseau et des rétroactions sont appliquées lorsqu'un produit réapparaît dans l'arbre. Dans la seconde situation, l'hypothèse de corrélation n'est pas faite, et c'est un arbre tronqué qui est considéré, dans lequel plusieurs instances d'un même produit peuvent apparaître et ne seront donc pas corrélées. Ce qui résulte de cette première analyse est que cette hypothèse est relativement peu sensible. Avec un critère de troncature de l'arbre à 1% des impacts totaux pour le cas non corrélé, les médianes des rapports des coefficients de variation pour l'ensemble des processus unitaires (du cas corrélé sur le cas non corrélé) sont de 1.03 pour les changements climatiques, 1.05 pour la santé humaine, 1.19 pour la qualité des écosystèmes, 1.07 pour les ressources et 1.05 pour le prélèvement des eaux. Ces résultats montrent de plus que les incertitudes sont relativement plus importantes dans le cas avec l'hypothèse de corrélation intradonnées.

Dans la seconde partie de l'étude, l'influence de la corrélation intrasystème dans la propagation des incertitudes est considérée. Cette analyse intervient dans le contexte particulier de systèmes de produit construits à partir de données agrégées. C'est généralement le cas des industries lorsqu'elles font appel à des outils ACV simplifiés pour des raisons de simplicité et

de rapidité. Dans cette analyse, les variances et les covariances des différents impacts agrégés intervenant dans les systèmes sont utilisées. Dans cette étude, la variance d'un système complet sera calculée d'une part en considérant les différents impacts indépendants, et d'autre part en considérant leurs corrélations qui sont tirées des matrices de variances-covariances calculées à partir des échantillons de simulations de Monte-Carlo. Ainsi, la propagation des incertitudes dans les deux cas est comparée. Plus précisément, la part des covariances dans le calcul de la variance totale est observée et représente 20% pour les changements climatiques, 27% pour la santé humaine, 31% pour la qualité des écosystèmes, 32% pour les ressources et 6% pour le prélèvement des eaux.

Dans la troisième partie de l'étude, une analyse des incertitudes lors de la comparaison de deux systèmes de produits qui ne sont pas considérés comme indépendants est réalisée. En effet, tout comme la corrélation précédente, les différents produits agrégés composant les deux systèmes ne sont pas nécessairement indépendants. Leurs corrélations sont alors calculées comme précédemment, et les variances correspondant aux impacts de la différence des deux systèmes en considérant ces corrélations sont calculées. Ainsi, il est extrait les corrélations entre les impacts des deux systèmes qui sont de 23% pour les changements climatiques, 24% pour la santé humaine, 55% pour la qualité des écosystèmes, 68% pour les ressources et 56% pour le prélèvement des eaux. Ces corrélations peuvent diminuer lorsque les utilisateurs des logiciels ACV décident de définir les mêmes produits apparaissant dans les deux systèmes indépendants.

Finalement, une preuve de concept est réalisée pendant l'étude. Elle prend la forme d'un outil développé avec le langage de programmation python. Cet outil permet de construire deux systèmes de produits agrégés en précisant si les impacts d'un même produit apparaissant dans les deux systèmes sont indépendants ou non. De plus, il est possible de préciser si les impacts considérés sont considérés avec ou sans l'hypothèse de corrélation intradonnées. Ensuite, l'outil permet de calculer très rapidement les impacts des deux systèmes complets, ainsi que les incertitudes associées à ces impacts. Les informations d'incertitudes tiennent compte de la corrélation intrasystème et sont données sous forme de barres d'erreurs et de tableaux avec les informations statistiques (valeur déterministe, mode, écart type, coefficient de variation). De plus, un graphique permet de comparer les impacts des deux systèmes de produits tout en considérant la corrélation intrasystème et la corrélation intersystème.

7.2 Recommandations

La synthèse précédente montre donc qu'il est possible d'intégrer des informations d'incertitudes aux résultats fournis par les logiciels ACV simplifiés tout en répondant aux exigences

du monde industriel. La considération des incertitudes lors de l'évaluation des impacts en analyse de cycle de vie est primordiale et cette pratique a tout intérêt à se populariser pour améliorer la crédibilité des conclusions tirées. La méthode utilisée dans les travaux consiste à calculer par des simulations de Monte-Carlo les informations d'incertitudes associées aux impacts pour ensuite les utiliser directement de manière analytique dans les logiciels. Cette méthode nécessite de refaire les simulations lorsque les méthodes d'impacts sont changées ou encore lorsque la base de données évolue. Ce point peut être considéré comme une faiblesse, cependant il permet une certaine flexibilité. En effet, cette méthode peut donc s'appliquer à n'importe quelle base de données (même spécifique à une entreprise) et n'importe quelle méthode d'impact, de plus, si les données d'incertitudes ont besoin d'être affinées, la taille des échantillons des simulations de Monte-Carlo peut être augmentée, cela ne ralentira pas le calcul des impacts dans les logiciels ACV. Finalement, certaines hypothèses sur l'allure des lois de distributions affaiblissent la pertinence des résultats d'incertitude présentés, par exemple l'hypothèse de normalité des impacts de la différence de deux systèmes, ou encore l'hypothèse de log-normalité des impacts du cycle de vie agrégés d'un système de produit. Ces approximations peuvent être évitées si un plus grand nombre de données sont stockées et une autre méthode utilisée pour le calcul des impacts comme présenté dans la section Discussion.

RÉFÉRENCES

W. B. Arthur, Y. M. Ermoliev, et Y. M. Kaniovski, “Path-dependent processes and the emergence of macro-structure”, *European Journal of Operational Research*, vol. 30, no. 3, pp. 294–303, 1987.

L. Basson et J. G. Petrie, “An integrated approach for the consideration of uncertainty in decision making supported by life cycle assessment”, *Environmental Modelling and Software*, vol. 22, no. 2, pp. 167–176, 2007.

C. R. Bojacá et E. Schrevens, “Parameter uncertainty in lca : stochastic sampling under correlation”, *The International Journal of Life Cycle Assessment*, vol. 15, no. 3, pp. 238–246, 2010.

G. Bourgault, P. Lesage, et R. Samson, “Systematic disaggregation : a hybrid lci computation algorithm enhancing interpretation phase in lca”, *The International Journal of Life Cycle Assessment*, vol. 17, no. 6, pp. 774–786, 2012.

A. Ciroth, G. Fleischer, et J. Steinbach, “Uncertainty calculation in life cycle assessments”, *The International Journal of Life Cycle Assessment*, vol. 9, no. 4, pp. 216–226, 2004.

R. Frischknecht, N. Jungbluth, H.-J. Althaus, R. Hischier, G. Doka, R. Dones, T. Heck, S. Hellweg, G. Wernet, et T. Nemecek, “Overview and methodology. data v2. 0 (2007). ecoinvent report no. 1”, Ecoinvent Centre, Swiss Federal Laboratories for Materials Testing and Research (EMPA), Duebendorf (Switzerland), Report, 2007.

M. Goedkoop et R. Spriensma, “The eco-indicator99 : A damage oriented method for life cycle impact assessment : Methodology report”, 2001.

M. Goedkoop, R. Heijungs, M. Huijbregts, A. De Schryver, J. Struijs, et R. van Zelm, “Recipe 2008”, *A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level*, vol. 1, 2009.

R. Heijungs, “Sensitivity coefficients for matrix-based lca”, *The International Journal of Life Cycle Assessment*, vol. 15, no. 5, pp. 511–520, 2010.

R. Heijungs et M. Lenzen, “Error propagation methods for lca—a comparison”, *The International Journal of Life Cycle Assessment*, vol. 19, no. 7, pp. 1445–1461, 2014.

R. Heijungs et S. Suh, *The computational structure of life cycle assessment*. Springer Science and Business Media, 2002, vol. 11.

R. Heijungs, S. Suh, et R. Kleijn, "Numerical approaches to life cycle interpretation-the case of the ecoinvent'96 database (10 pp)", *The International Journal of Life Cycle Assessment*, vol. 10, no. 2, pp. 103–112, 2005.

J. Hong, S. Shaked, R. K. Rosenbaum, et O. Jolliet, "Analytical uncertainty propagation in life cycle inventory and impact assessment : application to an automobile front panel", *The International Journal of Life Cycle Assessment*, vol. 15, no. 5, pp. 499–510, 2010.

M. A. Huijbregts, "Application of uncertainty and variability in lca", *The International Journal of Life Cycle Assessment*, vol. 3, no. 5, pp. 273–280, 1998.

H. Imbeault-Tétreault, O. Jolliet, L. Deschênes, et R. K. Rosenbaum, "Analytical propagation of uncertainty in life cycle assessment using matrix formulation", *Journal of Industrial Ecology*, vol. 17, no. 4, pp. 485–492, 2013.

ISO, "14040 : Environmental management–life cycle assessment–principles and framework", *International Organization for Standardization*, 2006.

——, "14044 : environmental management—life cycle assessment—requirements and guidelines", *International Organization for Standardization*, 2006.

O. Jolliet, M. Margni, R. Charles, S. Humbert, J. Payet, G. Rebitzer, et R. Rosenbaum, "Impact 2002+ : A new life cycle impact assessment methodology", *The International Journal of Life Cycle Assessment*, vol. 8, no. 6, pp. 324–330, 2003.

E. Limpert, W. A. Stahel, et M. Abbt, "Log-normal distributions across the sciences : Keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal : That is the question", *BioScience*, vol. 51, no. 5, pp. 341–352, 2001.

S. M. Lloyd et R. Ries, "Characterizing, propagating, and analyzing uncertainty in life-cycle assessment : A survey of quantitative approaches", *Journal of Industrial Ecology*, vol. 11, no. 1, pp. 161–179, 2007.

S.-C. Lo, H.-w. Ma, et S.-L. Lo, "Quantifying and reducing uncertainty in life cycle assessment using the bayesian monte carlo method", *Science of the Total Environment*, vol. 340, no. 1, pp. 23–33, 2005.

M. MacLeod, A. J. Fraser, et D. Mackay, “Evaluating and expressing the propagation of uncertainty in chemical fate and bioaccumulation models”, *Environmental Toxicology and Chemistry*, vol. 21, no. 4, pp. 700–709, 2002.

S. Muller, P. Lesage, A. Ciroth, C. Mutel, B. Weidema, et R. Samson, “The application of the pedigree approach to the distributions foreseen in ecoinvent v3”, *The International Journal of Life Cycle Assessment*, pp. 1–11, 2014.

T. J. Osler, “Taylor’s series generalized for fractional derivatives and applications”, *SIAM Journal on Mathematical Analysis*, vol. 2, no. 1, pp. 37–48, 1971.

G. Rebitzer, T. Ekvall, R. Frischknecht, D. Hunkeler, G. Norris, T. Rydberg, W. P. Schmidt, S. Suh, B. P. Weidema, et D. W. Pennington, “Life cycle assessment part 1 : framework, goal and scope definition, inventory analysis, and applications”, *Environment international*, vol. 30, no. 5, pp. 701–20, 2004.

G. Rice, R. Clift, et R. Burns, “Comparison of currently available european lca software”, *The International Journal of Life Cycle Assessment*, vol. 2, no. 1, pp. 53–59, 1997.

G. Ritschard, “Analyse quantitative des relations de causalité”.

S. M. Ross, *Introduction to Probability Models*, 8e éd., 2003.

S. Suh et R. Heijungs, “Power series expansion and structural analysis for life cycle assessment”, *The International Journal of Life Cycle Assessment*, vol. 12, no. 6, pp. 381–390, 2007.

S. Suh et G. Huppes, “Methods for life cycle inventory of a product”, *Journal of Cleaner Production*, vol. 13, no. 7, pp. 687–697, 2005.

W. Wei, P. Larrey-Lassalle, T. Faure, N. Dumoulin, P. Roux, et J.-D. Mathias, “How to conduct a proper sensitivity analysis in life cycle assessment : Taking into account correlations within lci data and interactions within the lca calculation model”, *Environmental Science and Technology*, 2015.

Z. Xu et S. N. Srihari, “Bayesian network structure learning using causality”, dans *Pattern Recognition (ICPR), 2014 22nd International Conference on.* IEEE, Conference Proceedings, pp. 3546–3551.

ANNEXE A BASE DE DONNÉES ECOINVENT

Dans cette annexe, quelques précisions utiles à l'étude sont données à propos de la base de donnée utilisée. Cette base de données est la version v2.2 de **ecoinvent**. Elle contient 4161 processus avec les 4161 flux économiques associés, ainsi que 1615 flux élémentaires.

Ces flux sont réparti dans deux matrices, la matrice technologique qui comprend les flux économiques (notée **A**) et la matrice intervention qui contient les flux élémentaires (notée **B**). La majorité des valeurs de ces matrices sont nulles (les matrices sont dites creuses). Plus de 99% des valeurs de la matrice technologique sont nulles, les autres valeurs sont des variables aléatoires décrites dans le tableau A.1. Plus de 98% des éléments la la matrice intervention sont nulles, les valeurs non nulles sont des variables aléatoires aussi décrites dans le tableau A.1.

Tableau A.1 Répartition des variables aléatoires dans les matrices technologique et intervention.

Distribution	Matrice technologique		Matrice intervention	
	Nbre	%	Nbre	%
Log-normale	39035	89.5%	55973	60%
Normale	27	0.06%	16	0.02%
Triangulaire	8	0.02%	0	0%
Inconnue	4546	10.5%	37098	39.9%

Les variables aléatoires suivant la loi log-normale composent la majorité de ces matrices. Les histogrammes de la Figure A.1 représentent la répartition des paramètres d'incertitudes (σ) des variables aléatoires dans chacune des matrices. Cela permet de donner une idée de l'incertitude globale de ces matrices.

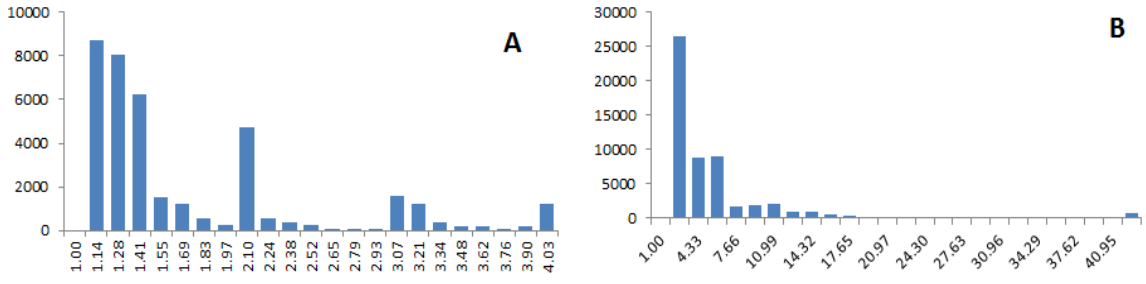


Figure A.1 Histogrammes représentant la répartition des valeurs des paramètres d'incertitudes des variables aléatoires Log-normales dans **A** et **B**.

Certaines de ces variables aléatoires semblaient erronées. Il semblerait que certaines variables aient été renseignées comme des variables aléatoires suivant des lois normales avec des paramètres d'ordre de grandeur largement différents. Ce qui n'est pas envisageable pour une variable aléatoire suivant une loi Normale. Le tableau A.2 décrit les flux qui semblent erronés ainsi que les variables qui expliqueraient cette erreur.

Tableau A.2 Variables aléatoires considérées comme normales, mais qui semblent être des log-normales.

Flux économique	Valeur déterministe	Paramètre d'incertitude
Bicycle, at regional storage/RER/I U Section bar extrusion, aluminium/RER U	-3.77E-6	0.78
Electric scooter, at regional storage/RER/I U Copper, at regional storage/RER U	-6.76E-7	0.79
Maintenance, bicycle/CH/I U Section bar extrusion, aluminium/RER U	-3.77E-7	0.79
Maintenance, electric vehicle, LiMn2O4/RER/I U Copper, at regional storage/RER U	-3.00E-7	0.64
Maintenance, passenger car, electric, LiMn2O4, city car/RER/I U Copper, at regional storage/RER U	-1.45E-7	0.64
Passenger car, diesel, EURO5, city car, at plant/RER/I U Copper, at regional storage/RER U	-4.21E-6	0.63
Passenger car, electric, LiMn2O4, at plant/RER/I U Copper, at regional storage/RER U	-1.01E-5	0.64
Passenger car, electric, LiMn2O4, city car, at plant/RER/I U Copper, at regional storage/RER U	-4.68E-6	0.64
Flux élémentaires		
Kenaf fibres, at farm/IN U Soil, Oils, unspecified, agricultural	3.17E-4	0.85
Kenaf stalks, from fibre production, at farm/IN U Soil, Oils, unspecified, agricultural	3.17E-4	0.85
Operation, electric scooter, certified electricity/CH U Air, Nickel, (unspecified)	7.85E-13	2.54
Operation, electric scooter/CH U Air, Nickel, (unspecified)	7.85E-13	2.54
Operation, passenger car, electric, LiMn2O4, city car/CH U Air, Chromium, (unspecified)	3.93E-12	2.64
Operation, passenger car, electric, LiMn2O4/CH U Air, Chromium, (unspecified)	8.25E-12	2.64

ANNEXE B LOG-NORMALITÉ DES IMPACTS

Comme il est expliqué dans la section 4.1.1, des simulations de Monte-Carlo sont réalisées afin d'obtenir des échantillons des variables aléatoires correspondant aux impacts des systèmes élémentaires de tous les processus de la base de données **ecoinvent v2.2**. Ensuite, lors de la construction des systèmes de matrices correspondant au cas désagrégé, les données d'incertitudes correspondant aux impacts de ces systèmes de produits sont utilisés. Les étapes sont détaillées dans la section 4.1.2. Finalement, lors de la simulation de Monte-Carlo dans le cas non corrélé, les impacts agrégés doivent être tirés aléatoirement selon une certaine loi. Une étude statistique sur les échantillons nous a permis de faire l'hypothèse que les résultats d'impact agrégés suivaient des loi Log-normales.

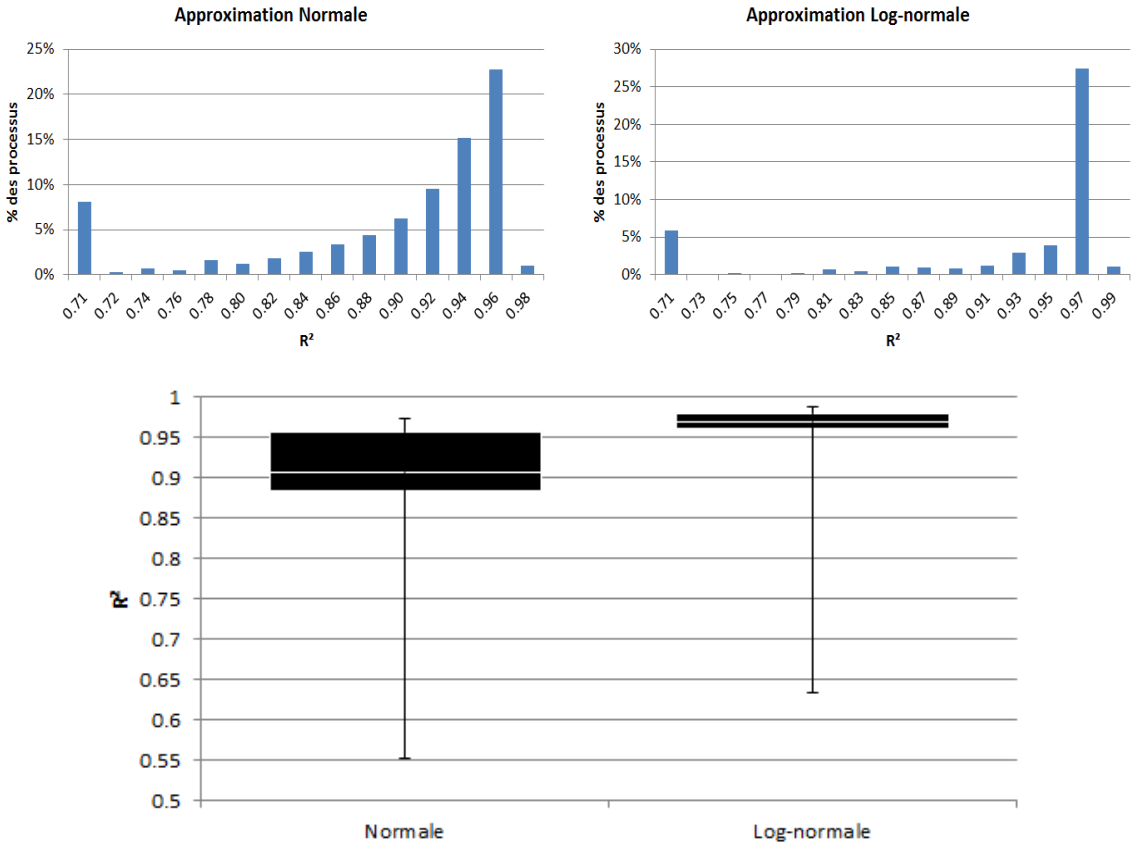


Figure B.1 Histogrammes et box-plot représentant les R^2 pour l'ensemble des processus par rapport à des approximations Normale et Log-normale.

Une première observation d'un certain nombre d'histogrammes ont permis de remarques que la distribution des impacts ressemblait à celle d'une loi Log-normal ou d'une loi Normale.

Ensuite, une étude plus précise a été réalisée. Les paramètres μ et σ des lois Normales et Log-normales ont été estimés par la méthode des moments pour chacun des échantillons qui correspondent respectivement aux impacts de tous les systèmes de processus élémentaires de la base de données. Ensuite, afin d'avoir une idée de l'erreur réalisée lors de l'hypothèse de normalité ou de log-normalité, les R^2 ont été calculés.

Cette étude nous a permis de savoir si il était préférable de faire l'hypothèse que les distributions étaient plutôt Normales ou Log-normales, mais aussi de connaître la validité de cette hypothèse. Finalement, il est aussi possible d'identifier les processus élémentaires pour lesquels l'hypothèse n'est pas valide, et de s'assurer lors de l'analyse des résultats de l'étude que ces processus ne sont pas prépondérants dans les calculs réalisés.

Les calculs ont seulement été faits pour la catégorie de dommage, celle des changements climatiques. Cependant, étant donné que la matrice des facteurs de caractérisation est n'est pas aléatoire, on considérera les résultats généralisables aux autres catégories de dommages.

Les boxplots et les histogrammes de la figure B.1 montrent que les l'approximation Log-normale est nettement mieux que l'approximation Normale. De plus, on peut voir que plus de 75% des variables aléatoires ont un R^2 supérieur à 95%, ce qui nous permet d'avoir une idée de la validité de l'hypothèse de log-normalité pour les impacts. De plus, pour 100% des processus, le R^2 correspondant à l'approximation Log-normale est plus grand que celui de l'approximation Normale.

ANNEXE C NORMALITÉ DE LA DIFFÉRENCE D'IMPACTS

A présent dans cette annexe, il va être expliqué l'hypothèse de normalité de la section 4.3, dans laquelle on suppose que la différence entre les impacts de deux systèmes de produits suit une loi normale. En réalité, l'hypothèse qui est faite pour obtenir les résultats et moins forte que cela. En effet, le résultat ne porte que sur la probabilité que la différence des impacts soit supérieure à 0 (ou le contraire) en supposant la distribution normale.

Tout d'abord, la moyenne et la variance des impacts sont calculées comme présenté dans la section 4.3, puis la distribution des impacts de la différence des deux systèmes est supposée normale. Ensuite, les paramètres de la loi sont estimés grâce à la méthode des moments. Enfin, la probabilité que ces impacts soient négatifs (ou positifs) est calculée. Cela nous permet de savoir à quel point un système est meilleur que l'autre.

Afin de renseigner la validité de l'hypothèse de normalité, un échantillon de 13862 couples de différences. Afin d'avoir un échantillon représentatif de la base de données, un processus a été choisi dans chacune des catégories selon le découpage classique des logiciels ACV. Cela a permis d'extraire 167 processus relativement représentatifs de la totalité de la base de données. Ensuite, toutes les combinaisons possibles des différences ont été prises pour déterminer l'échantillon. Cela donne la combinaison de 2 parmi 167 : 13862. Ensuite, les impacts de ces processus sont tous normalisés par leur moyenne afin de passer outre les différents ordres de grandeur liés aux différences d'unités par exemple.

Par la suite, pour chaque couple, un échantillon de tirages de l'impact de la différence a été construit en utilisant les résultats de la simulation de Monte-Carlo de la section 4.1.1. Puis une approximation normale de la distribution de ces impacts est déterminée en utilisant la méthode des moments. Enfin, il est d'abord calculé le R^2 correspondant à l'approximation de la distribution de l'échantillon par la loi Normale, pour savoir s'il est acceptable de prétendre que la différence de deux impacts suit une loi Normale. Et finalement, la probabilité que la différence soit inférieure à 0 est comparée dans les deux cas (avec l'échantillon, et avec l'approximation normale). Dans le cas de l'échantillon, la probabilité recherchée correspond simplement au nombre de résultats négatifs divisé par la taille de l'échantillon. Dans le cas de l'approximation normale, la probabilité recherchée vaut systématiquement 0.5 car les impacts des deux processus sont centrés grâce à leurs moyenne et la loi Normale est symétrique.

$$\mu = \mathrm{E}\Big(\frac{\mathrm{Impact}_1}{\mathrm{E}(\mathrm{Impact}_1)} - \frac{\mathrm{Impact}_2}{\mathrm{E}(\mathrm{Impact}_2)}\Big) = \mathrm{E}\Big(\frac{\mathrm{Impact}_1}{\mathrm{E}(\mathrm{Impact}_1)}\Big) - \mathrm{E}\Big(\frac{\mathrm{Impact}_2}{\mathrm{E}(\mathrm{Impact}_2)}\Big)$$
$$= \frac{\mathrm{E}(\mathrm{Impact}_1)}{\mathrm{E}(\mathrm{Impact}_1)} - \frac{\mathrm{E}(\mathrm{Impact}_2)}{\mathrm{E}(\mathrm{Impact}_2)} = 1 - 1 = 0$$

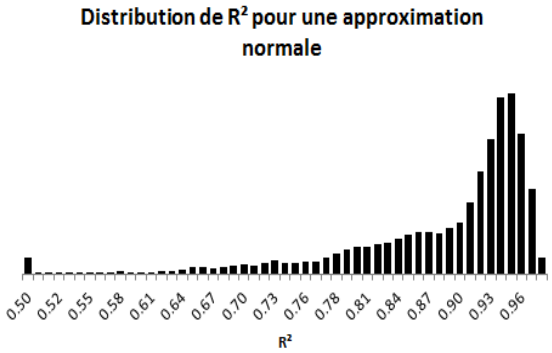


Figure C.1 Distribution des R^2 pour une approximation Normale de la différence de deux impacts.

La figure C.1 montre qu'une approximation Normale correspond pour la majeure partie des couples de processus. Cependant, on peut remarquer qu'une partie non négligeable des couples de processus ont un R^2 relativement faible pour l'approximation Normale. C'est pourquoi le résultat qui nous intéresse directement est aussi étudié : la probabilité que la différence entre les impacts de deux processus soit négative. Cette probabilité est comparée à la même probabilité en supposant que la distribution de la différence est Normale. Ces résultats sont représentés dans l'histogramme de la figure C.2.

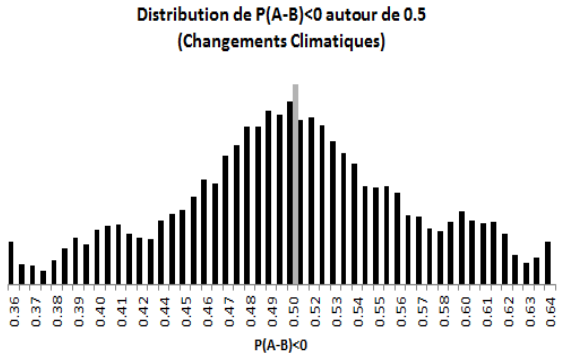


Figure C.2 Distribution des impacts de la différence de deux systèmes normalisés.

On peut donc y voir que les résultats tirés des échantillons de Monte-Carlo ne s'écartent pas significativement di cas avec l'hypothèse de normalité. Il est tout de même possible de trouver des couples de processus pour lesquels l'approximation donne un résultat relativement loin de l'échantillon. Ces cas isolés ont été observés plus en détail. Ces cas peuvent être catégorisés en deux familles distinctes composées chacunes de quelques disaines de cas (partie négligeable devant les 13862 couples). Les histogrammes de la figure C.3 illustrent les formes des distributions des cas particuliers.

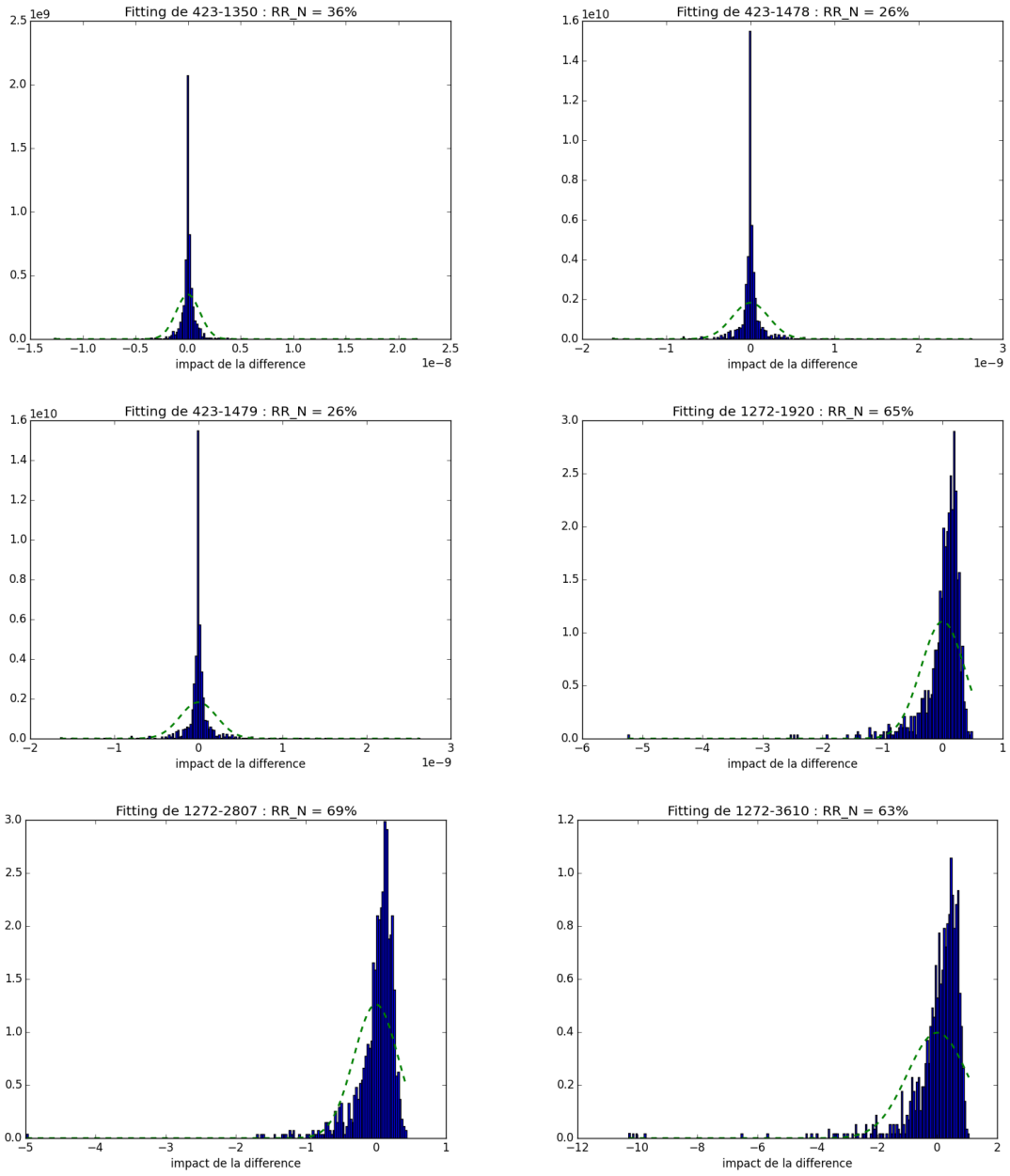


Figure C.3 Exemples de cas particuliers avec des R^2 faibles pour l'approximation normale.

ANNEXE D PRÉCISIONS SUR LA CORRÉLATION INTRASYSTÈME

Définition théorique.

Covariance :

La covariance entre deux variables aléatoires X et Y est la valeur de l'espérance du produit des écarts des variables avec leur espérance respective. Cette valeur quantifie la correspondance entre les variations des deux variables. Si les variables évoluent de manières similaires, la covariance sera positive. Et si elles évoluent de manières opposées, leur covariance sera négative. Finalement, si les variables évoluent de manières indépendantes, la valeur de la covariance sera nulle (attention, la réciproque est fausse : contres exemples).

Il faut faire attention à bien garder en tête que cette mesure ne rend compte que de la correspondance des variations et non de la dépendance/causalité d'une variable à une autre comme le montre (Ritschard) dans ses études sur la causalité.

$$\mathrm{Cov}(X,Y) = \mathrm{E}\Big[\Big(X - \mathrm{E}(X)\Big)\Big(Y - \mathrm{E}(Y)\Big)\Big] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X)$$

Corrélation :

La corrélation est le rapport de la covariance de deux variables sur le produit de leurs écarts types. La corrélation est donc une valeur sans dimension et comprise entre -1 et 1.

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \qquad \begin{cases} \rho = -1 \rightarrow \text{ corrélation inverse} \\ \rho = 0 \rightarrow \text{ corrélation nulle (attention il n'y a pas indépendance)} \\ \rho = -1 \rightarrow \text{ corrélation parfaite} \end{cases}$$

Propriétés de calculs

La covariance est une forme bilinéaire symétrique, cela implique les propriétés de calcul suivantes :

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) \quad \text{(symétrie)}$$

$$\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y) \quad \text{(bilinéarité)}$$

$$\mathrm{Var}(aX) = \mathrm{Cov}(aX, aX) = a^2\mathrm{Var}(X)$$

$$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y) \quad \text{(bilinéarité)}$$

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Cov}(X_1 + X_2, X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$$

$$\mathrm{Cov}(F_1, F_2) = \mathrm{Cov}\Big(\sum_i a_i X_i, \sum_j b_j Y_j\Big) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j) \quad \text{(bilinéarité généralisée)}$$

$$\mathrm{Var}(F_1) = \mathrm{Cov}\Big(\sum_i a_i X_i, \sum_j a_j X_j\Big) = \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j)$$

Exemples et représentations

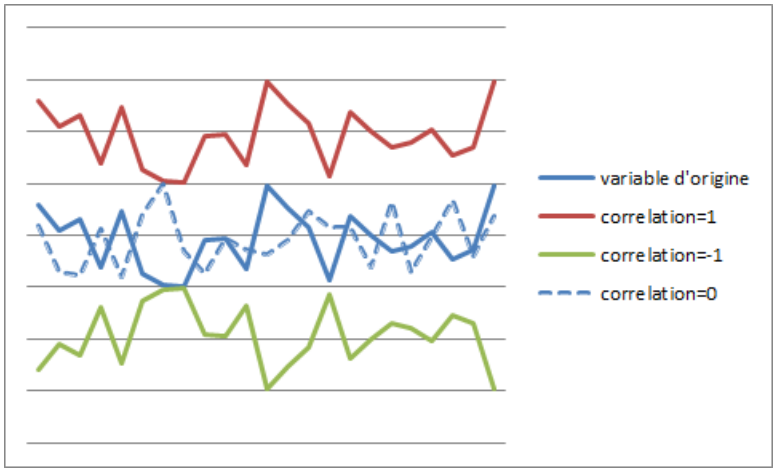


Figure D.1 Les différents cas extrêmes de corrélation

- En bleu continue : Une série de valeurs aléatoires (cela pourrait correspondre aux valeurs des impacts d'un processus dans une catégorie d'impact après une simulation de Monte-Carlo).
- En rouge continue : La série de valeurs est parfaitement corrélée avec la série d'origine. Nous remarquons que les variations de l'une et de l'autre sont identiques. Par exemple, cela pourrait arriver pour deux grid-mix de pays différents, mais qui utiliseraient exactement les mêmes processus en quantité similaires.
- En vert continue : La série de valeurs est en corrélation inverse avec la série d'origine. Nous remarquons que les variations de l'une et de l'autre sont parfaitement opposées. Par exemple, cette situation pourrait correspondre au cas d'un recyclage dont le seul impact proviendrait de la matière recyclée, et qui aurait un impact parfaitement opposé à celui correspondant au flux de la matière en question.

– En bleu pointillée : La série a été générée indépendamment de la série d'origine. Nous remarquons que les variations de l'une ne semblent pas liées aux variations de l'autre. Cette situation correspondrait à celle de deux processus parfaitement indépendants. Autrement dit, ils n'auraient aucun processus en commun dans leurs arbres respectifs.

Application à la corrélation entre les processus (intrasystème)

Dans notre étude, nous avons généré des simulations de Monte-Carlo pour chacun des processus de la base de données. Cela nous a permis d'obtenir une distribution empirique (histogramme), que nous avons associée à une distribution log-normale correspondante et donc aux paramètres de la loi : μ, σ. Ces deux valeurs nous suffisent pour avoir accès à l'impact et à son incertitude (variance).

À présent, nous souhaitons utiliser ces informations pour calculer l'impact et l'incertitude d'un système de processus (une combinaison linéaire de processus). Prenons le système suivant :

– X_1 : Un premier transport de l'usine à l'entrepôt (de 3pkm par exemple)
– X_2 : Un deuxième transport de l'entrepôt au point de vente (de 4pkm par exemple)

Nous connaissons les distributions des impacts de ces deux processus : Log-normale(μ_i, σ_i). Nous cherchons à connaître l'impact avec incertitude (variance) de ce système : $F = 3X_1 + 4X_2$

En considérant nos deux processus complétement indépendants, on a les résultats suivants :

$$\text{impact}(F) = 3\ \text{impact}(X_1) + 4\ \text{impact}(X_2) \quad \text{avec} \quad \text{impact}(X_i) = \text{mode}(X_i) = e^{\mu_i - \sigma_i^2}$$

$$\text{Var}(F) = \text{Var}(3X_1 + 4X_2) = 9\text{Var}(X_1) + 16\text{Var}(X_2) \quad \text{avec} \quad \text{Var}(X_i) = (e^{\sigma_i^2} - 1)e^{2\mu_i + \sigma_i^2}$$

Cependant les processus que nous avons choisis ne sont pas indépendants. Cela provient du fait que leurs arbres/réseaux font très certainement appel aux mêmes processus. Plus précisément, les impacts des processus X_1 et X_2 ont été obtenu à partir des flux présents dans la matrice technologique($\mathbf{A}$) et la matrice d'intervention($\mathbf{B}$).

$$X_1 = F_1(A, B) = C_F^\top B A^{-1} f_1 \qquad f_1 \text{ correspond au vecteur de demande de } X_1$$

$$X_2 = F_2(A, B) = C_F^\top B A^{-1} f_2 \qquad f_2 \text{ correspond au vecteur de demande de } X_2$$

Nous voyons bien ici que les impacts de $\mathbf{X_1}$ et de $\mathbf{X_2}$ sont des fonctions (complexes) des

mêmes variables aléatoires qui sont les flux économiques et élémentaires : A_{ij} et B_{ij}. Leur covariance n'a alors aucune raison d'être nulle.

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(F_1(A, B), F_2(A, B)) = \mathrm{Cov}(C_F^\top B A^{-1} * f_1, C_F^\top B A^{-1} f_2)$$

Pour obtenir le résultat des covariances, on peut appliquer deux méthodes :

- Par le calcul (comme dans l'exemple). Cette méthode consisterait à décomposer $F_1(\mathbf{A}, \mathbf{B})$ et $F_2(\mathbf{A}, \mathbf{B})$, puis utiliser les propriétés de calcul de la covariance vue dans la section D. Cependant, nous conviendrons que cette méthode sera extrêmement complexe.
- Par des tirages aléatoires. Tout comme la méthode de Monte-Carlo, nous tirons successivement des valeurs pour tous les flux selon leurs distributions. Nous calculons X_1 etX_2 à chaque itération. Nous pouvons alors appliquer la fonction de covariance aux deux vecteurs des données de X_1 et X_2 avec **Matlab**, **R** ou même **Excel**. Cependant, il faut faire attention de ne pas tirer les valeurs des flux utilisés pour calculer X_1 et X_2 de manière indépendante.

C'est la 2$^{\mathrm{e}}$ méthode qui sera retenue dans notre étude. Car nous avons déjà à notre disposition les vecteurs aléatoires X_i. Cependant, nous devrons absolument prendre les résultats obtenus à partir du cas corrélé. Car dans le cas non corrélé, les calculs des X_i sont fait de manières indépendantes (matrices $\mathbf{A}$ et $\mathbf{B}$ différente pour chaque X_i).

Nous avons alors calculé la covariance entre ces deux processus grâce aux simulations de Monte-Carlo du cas corrélé. Notons cette covariance :c_{12}. En considérant la covariance donnée par la simulation de Monte-Carlo, on aura les résultats suivants :

$$\mathrm{impact}(F) = 3\ \mathrm{impact}(X_1) + 4\ \mathrm{impact}(X_2)$$

$$\mathrm{Var}(F) = \mathrm{Var}(3X_1 + 4X_2) = 9\ \mathrm{Var}(X_1) + 16\ \mathrm{Var}(X_2) + 12c_{12}$$

Généralisation à la corrélation de deux systèmes de processus.

La notion de corrélation entre deux processus peut se généraliser à un plus grand nombre de processus. La bilinéarité généralisée de la COVARIANCE nous permet de calculer la VARIANCE de n'importe quelle combinaison de processus.

$$\mathrm{Var}(\sum_i a_i X_i) = \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$$

Dans ces expressions, on peut voir que la matrice de variance-covariance ($\mathrm{Cov}(X_i, X_j)$) nous suffit pour calculer la variance d'un système de processus, et la covariance/corrélation entre deux systèmes de processus.