



HAL
open science

EVALUATION ENVIRONNEMENTALE COUPLEE A L'ANALYSE MULTIDIMENSIONNELLE DES DONNEES POUR L'ECONOMIE CIRCULAIRE

N Prioux, R Ouaret, Z. Chen, G Hetreux, J.-P Belaud

► **To cite this version:**

N Prioux, R Ouaret, Z. Chen, G Hetreux, J.-P Belaud. EVALUATION ENVIRONNEMENTALE COUPLEE A L'ANALYSE MULTIDIMENSIONNELLE DES DONNEES POUR L'ECONOMIE CIRCULAIRE. 13ème CONFERENCE INTERNATIONALE DE MODELISATION, OPTIMISATION ET SIMULATION (MOSIM2020), 12-14 Nov 2020, AGADIR, Maroc, Nov 2020, AGADIR (virtual), Maroc. hal-03192823

HAL Id: hal-03192823

<https://hal.science/hal-03192823>

Submitted on 8 Apr 2021

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

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

EVALUATION ENVIRONNEMENTALE COUPLEE A L'ANALYSE MULTIDIMENSIONNELLE DES DONNEES POUR L'ECONOMIE CIRCULAIRE

N. PRIoux, R. OUARET, Z. CHEN, G. HETREUX et J.-P. BELAUD
Laboratoire de Génie Chimique, Université de Toulouse, CNRS, Toulouse, France
nancy.prioux@ensiacet.fr

RESUME : *L'économie circulaire et ses différentes boucles de recirculation sont devenues ces dernières années un objet d'étude majeur, notamment dans le domaine de l'agriculture qui est un important pourvoyeur de déchets. De nombreuses recherches sont menées pour transformer les déchets lignocellulosiques de l'agriculture par des procédés « durables » c'est-à-dire économiquement viables, socialement acceptés et respectueux de l'environnement. Par la "pensée cycle de vie", il est possible d'évaluer ces impacts environnementaux potentiels. Cependant, ces analyses environnementales nécessitent en général un volume important de données spécifiques, dont la collecte peut être longue et fastidieuse, ou simplement impossible dans la pratique. D'autre part, les articles scientifiques décrivant les procédés de valorisation des coproduits de l'agriculture constituent une source de données intéressante mais peu exploitée. Dans cet article, une approche générale couplant la science des données et l'analyse environnementale est proposée. Composée de cinq étapes, cette approche est orientée vers une aide à la décision pour le chercheur ou l'ingénieur R&D lors d'une étude préliminaire. Elle est mise en œuvre dans le cadre de l'étude des procédés de prétraitements de la tige de maïs et de la paille de riz.*

MOTS-CLES : *Développement durable, Simulation, Analyse du cycle de vie, Science des données, Procédés de prétraitement, Déchets agricoles.*

1 INTRODUCTION ET CONTEXTE SCIENTIFIQUE

Il est admis aujourd'hui que nos modes de vie et de consommation ne peuvent conserver le schéma linéaire "extraire, fabriquer, utiliser et jeter" comme modèle principal. C'est pourquoi la Commission Européenne a proposé l'utilisation du modèle d'économie circulaire (EC) pour stimuler l'utilisation de modèles durables (Ellen Macarthur Foundation, 2015). En 2015, le gouvernement français a proposé la SNTEDD (Stratégie Nationale de Transition Ecologique pour le Développement Durable) qui est composée de neuf piliers dont l'EC. Selon l'Agence française de l'environnement et de la maîtrise de l'énergie (ADEME), l'EC prend en compte trois domaines d'action: (1) la consommation par la demande et le comportement du consommateur, (2) l'offre et les acteurs économiques dont l'écologie industrielle est une voie admise et prometteuse dès la conception initiale d'une zone territoriale (Belaud et al., 2019a) et (3) la gestion des déchets. Ces trois domaines d'action représentent l'ensemble du cycle de vie d'un produit, d'un service ou d'un procédé. Pour obtenir des modèles durables, la pensée *cycle de vie* peut contribuer à améliorer les performances environnementales et à optimiser les avantages économiques et sociaux. Plusieurs méthodes globales sont apparues pour concevoir des procédés de recyclage des déchets qui s'intègrent dans l'économie circulaire (Grimaud et al., 2017). L'agriculture est un domaine particulier dans lesquelles l'EC et la pensée « cycle de vie » se sont développées au cours des dernières décennies.

En France, les industries agroalimentaires produisent environ 2,6 millions de tonnes de déchets organiques par

an, un chiffre en constante augmentation (Barry, 2020). Cette tendance s'accompagne de l'augmentation prévue de la population mondiale. Par ailleurs, les activités humaines réduisent la quantité de terres disponibles pour l'agriculture, ce qui a inévitablement des répercussions sur les systèmes agricoles. Bien que la biomasse lignocellulosique soit une ressource renouvelable, les procédés de transformation se doivent d'être durables pour participer au développement durable. Au cours des trente dernières années, de nombreux procédés de prétraitement ont été étudiés et publiés (Davis et al., 2017). Cependant, il y a un manque de critères pour guider le choix entre tous ces procédés. Le recours à l'évaluation environnementale, économique et sociale dans un contexte d'EC est un bon moyen d'orienter ce choix. Ces évaluations nécessitent de nombreuses données, en particulier des données techniques sur les procédés potentiellement disponibles dans les articles scientifiques.

Sans les outils adaptés, il est cependant difficile de pouvoir traiter un grand nombre de données. C'est pourquoi avec l'arrivée du « Big Data » ou « données massives » de nombreux outils et méthodes ont été créés. Les évaluations environnementales telle que l'analyse du cycle de vie (ACV) bénéficieraient clairement de ces technologies pour analyser les données des procédés et les résultats de ces évaluations. Quelques études utilisent les données massives pour compléter l'ACV. J. Cooper et al. utilisent les données massives pour compléter les données d'arrière-plan (Cooper et al., 2013). La proposition DILCA (Data-intensive Life Cycle Assessment) exploite l'ingénierie des connaissances (IC) et intègre les développements technologiques qui peuvent modifier les résultats de l'ACV pour un produit donné au fil du temps (Bhinge et al., 2015)

Le but de cet article est de présenter une approche d'aide à l'analyse et à la comparaison de différents procédés de prétraitement de la biomasse lignocellulosique et de différentes biomasses dans le contexte de l'économie circulaire. Cette approche est prévue pour le chercheur ou l'ingénieur R&D qui voudrait faire un premier tri dans les procédés de prétraitement ou les biomasses grâce aux données contenues dans la littérature scientifique. Elle peut être déployée lors d'une phase d'étude préliminaire et servir d'aide à une première décision pour le développement d'un pilote de laboratoire ou semi-industriel. Après une introduction des différentes méthodes utilisées dans notre approche, le paragraphe 2 expose cette approche générale. Le paragraphe 3 traite de l'étude de différents prétraitements de la paille de riz et la tige de maïs puis, fournit les résultats et les interprétations.

2 APPROCHE METHODOLOGIQUE

2.1 Méthodes

L'approche couple deux domaines distincts que sont la science des données et l'analyse environnementale. De nombreuses données de nature différente et issues de sources hétérogènes sont nécessaires pour l'analyse environnementale. Par exemple, les données procédés sont issues du web public au travers des articles scientifiques et constituent une base de données non structurée alors que les données d'arrière-plan sont disponibles dans des bases de données structurées (publiques ou privées) comme EcoInvent (Frischknecht et al., 2005). La science des données va fournir des méthodes et technologies adaptées pour cette analyse. L'objectif principal de cette approche est d'analyser différentes voies technologiques des procédés et de fournir une aide à la décision par l'analyse des résultats.

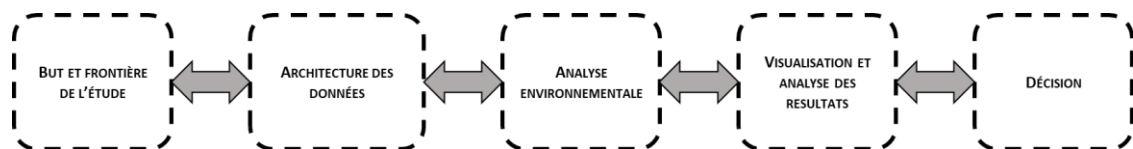


Figure 1: Schéma de l'approche générale

Le choix pour l'analyse environnementale s'est porté sur la méthode d'analyse du cycle de vie (ACV) (ISO 14040:2006, 2006). L'ACV comprend quatre étapes : la définition de l'objectif et du champ d'étude, l'inventaire du cycle de vie, l'évaluation des impacts du cycle de vie (EICV) et l'analyse et l'interprétation des résultats. Cette dernière étape de l'ACV est « délicate » pour des novices voire des experts. Il peut donc être profitable d'enrichir la méthode ACV pour aider le chercheur ou l'ingénieur à analyser et interpréter.

La littérature enseigne que les méthodes les plus appliquées pour l'analyse des résultats sont établies principalement sur la connaissance, l'expertise et la prise

de décision multicritère. Il existe quelques études récentes menées à l'aide d'algorithmes d'apprentissage automatique. Par exemple, (Kouchaki-Penchah et al., 2017) et (Zhao et al., 2019) ont combiné l'analyse par enveloppement des données (DEA) et l'ACV. Romeiko et al ont utilisé des modèles de régression arborescente accélérée pour identifier les principaux facteurs influents parmi les sols, le climat et les pratiques agricoles, qui déterminent l'hétérogénéité spatiale et temporelle des impacts environnementaux du cycle de vie (Romeiko et al., 2020).

Les méthodes d'apprentissage utilisées pour l'évaluation générale de la durabilité sont en général « supervisées ». Dans ce cas de figure, la prédiction (régression ou classification) se fait à partir des données « annotées » (Abdella et al., 2020). Par contre, dans de nombreux cas, les données ne sont pas « annotées » et l'objectif est d'extraire des classes de variables/échantillons présentant des caractéristiques communes. Pour mettre en évidence les similarités entre les différents procédés, les différentes biomasses ainsi que les similarités entre les impacts environnementaux, notre axe de travail repose sur l'utilisation de techniques de réduction de la dimension (DR). Plusieurs stratégies ont été explorées pour atteindre cet objectif en utilisant l'apprentissage « non supervisé » qui comprend la réduction de la dimension et le regroupement des données.

2.2 Approche générale

L'approche présentée dans ce papier repose sur une architecture générale prenant en compte les piliers de la durabilité (Belaud et al., 2019b). Cette approche est divisée en cinq grandes étapes (Figure 1). Chaque étape possède des sous-étapes et le passage de l'une à l'autre peut se faire dans un sens ou dans l'autre par des boucles de rétroaction. Il est recommandé d'itérer afin de consolider les résultats et donc les choix qui découlent de leurs interprétations.

Dans la première étape, les *but et frontière de l'étude* doivent être clairement définis. La pensée cycle de vie est recommandée. Cette pensée encourage une approche du « berceau à la tombe » voire « du berceau à la porte » si la logistique avale d'une chaîne de valeur est difficilement accessible. Les limites d'un système et l'unité fonctionnelle sont ensuite détaillées. Elles peuvent influencer grandement les évaluations. Par exemple, il faut préciser si la chaîne d'approvisionnement de la biomasse en amont est prise en compte. Une fois l'objectif et le champ d'application proprement définis, la chaîne d'approvisionnement, les technologies et les procédés de transformation doivent être décrits.

L'architecture des données est directement inspirée de la construction de l'architecture des données massives et comportent cinq sous étapes : (i) la collection et l'extraction des données, (ii) l'enrichissement et le stockage des données, (iii) le traitement des données, (iv) l'analyse des données (brutes) et (v) la visualisation des données (brutes). Cette étape permet de collecter, stocker et traiter toutes les données dont l'analyse environnementale aura besoin. Ces données sont divisées en deux groupes : les données procédés (dites de premier plan) qui sont extraites de la littérature scientifique et les données d'arrière-plan issues de bases de données libres ou commerciales comme EcoInvent. Cette étape (qui n'est pas documentée dans la section 3) utilise des outils de la science des données comme l'ingénierie des connaissances pour la collection et l'extraction des données.

Vient alors l'étape de l'analyse environnementale qui permet d'obtenir des impacts environnementaux à partir des données traitées issues de l'étape précédente. Ici, c'est l'ACV qui est sélectionnée. Une fois l'inventaire du cycle de vie (ICV) effectué, une ou plusieurs méthodes de calcul des impacts doivent être déterminées en accord avec la première étape qui intègre la nature de l'étude et du système. A la fin de cette étape, le résultat principal est une structure [procédés : biomasses : impacts] qui est difficile d'analyser.

L'étape de visualisation et d'analyse des résultats va apporter les méthodes issues de l'intelligence artificielle et plus précisément du « machine learning » pour aider à l'analyse de cette structure. A partir de la littérature statistique, sont combinées les techniques traditionnelles de réduction des dimensions et de classification non supervisée pour extraire les informations sur l'analyse des impacts du cycle de vie (AICV). Plus précisément, l'approche hybride est basée sur l'échelle multidimensionnelle (MDS) en utilisant la distance de Canberra et les k -moyennes. L'objectif est de rechercher des structures "cachées" dans les données multidimensionnelles et d'aider à interpréter la zone de points terminaux groupés de la matrice d'évaluation AICV. L'avantage de cette approche est que les méthodes basées sur les données nécessitent très peu de connaissances des procédés pour effectuer cette tâche. **Erreur ! Source du renvoi introuvable.** résume le traitement pour une structure [procédés : impacts] qui est exploitée dans l'étude présentée dans le paragraphe 3.

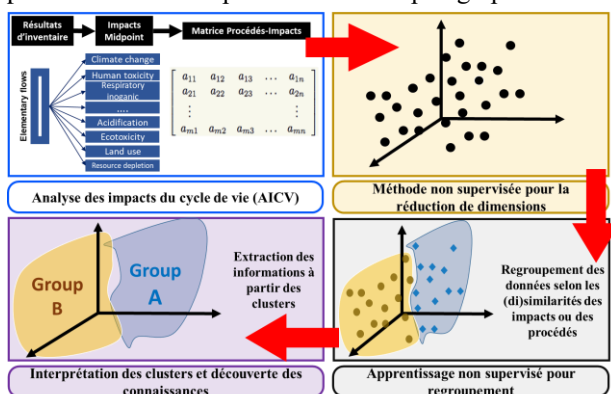


Figure 2 : Schéma du traitement des données pour l'AICV

La structure [procédés : biomasse : impacts] est d'abord utilisée comme entrée (matrice de similarité) pour tirer des enseignements des données qui sont les résultats de l'étape précédente. Deuxièmement, les techniques de DR projettent les données brutes du procédé dans un espace dimensionnel inférieur (2 ou 3). Le sens de dimension ici est celui de l'algèbre linéaire : la dimension de l'espace vectoriel sous-jacent des variables. Après une projection des données par une technique de DR, l'approche de regroupement est alors appliquée pour considérer les impacts, les biomasses similaires ainsi que dans les procédés au sein de l'espace dimensionnel inférieur. Enfin, l'utilisateur (expert) analyse les points regroupés en clusters pour les relier à des procédés/impacts/biomasses significatifs. Cette dernière étape d'affectation des clusters et d'extraction des informations est appelée « découverte de connaissances ».

La visualisation des clusters des données (résultats de l'étape précédente) va pouvoir aider le chercheur ou l'ingénieur R&D dans l'étape finale de décision. Cette décision peut être prise par un individu ou un groupe multidisciplinaire selon différentes techniques de prise de décision collaborative.

3 APPLICATION ET RESULTATS

Dans cette partie, les premiers résultats de notre approche sont présentés avec objet d'étude la comparaison de différents procédés de prétraitements de deux biomasses : la tige de maïs et la paille de riz. Ce paragraphe illustre la première étape de la méthode ie les *but et frontières de l'étude*. Cette étude examine quinze procédés provenant de quinze articles scientifiques. La Figure 3 **Erreur ! Source du renvoi introuvable.** illustre un de ces procédés de prétraitement. Le but de l'étude est d'aider le chercheur ou l'ingénieur R&D à choisir un procédé et/ou une biomasse pour la production de glucose. La pensée cycle de vie est du « berceau à la porte » c'est-à-dire que les frontières de l'étude sont de la biomasse à l'étape de l'hydrolyse enzymatique. La biomasse étant ici un déchet (voué à devenir un coproduit) de l'agriculture, elle comporte un impact nul – les impacts de la phase agricole sont imputés au produit final valorisé de l'agriculture (le maïs et le riz). La bioraffinerie est considérée relativement à proximité du champ, l'étape de transport étant alors négligeable. La fonction principale du système est la production du glucose. L'unité fonctionnelle est alors la « production de 1 kg de glucose » et tous les résultats sont exprimés en unité fonctionnelle.

Nous passons ici à la seconde étape l'architecture des données. Cette étape décrite ici succinctement est approfondie dans (Belaud et al., 2019b). La sélection des articles décrivant les procédés ayant comme intrant la tige de maïs ou la paille de riz est effectuée par des experts du domaine. Dix-huit articles sont sélectionnés dans un premier temps et extraits des bases de données scientifiques telles que Web of science et Science Direct. Les données des articles sont extraites de manière semi-automatique

dans une ontologie. L'ontologie structure les données procédés et assure un export en fichiers CSV alimentant un logiciel interne. Ce logiciel développé sur MS Excel réalise un premier « nettoyage » des données par la simulation des procédés pour calculer et vérifier le bilan matière.

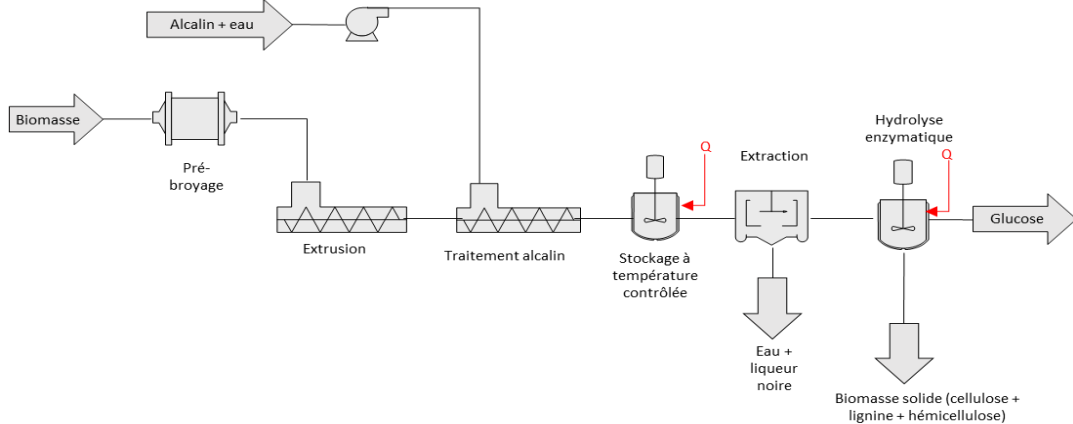


Figure 3 : Exemple d'un procédé de prétraitement étudié adapté de (Liu et al., 2013)

Ce nettoyage peut consister à des suppressions ou des modifications de valeurs. Par exemple, les données de 3 articles sont retirées de notre étude car ils présentent des incohérences ou de nombreuses données manquantes ce qui ne permet pas les vérifications par la simulation. L'étude analyse donc 15 procédés par 2 biomasses. L'évaluation environnementale utilise une méthode ACV (attributionnelle) : ReCiPe 2016, la base de données d'arrière-plan est EcoInvent v2.2 et les données de premier plan sont les données procédés « nettoyées ». L'évaluation environnementale conduit au calcul de 17 impacts dit « mid-point ». L'outil Excel a été validé à l'aide du simulateur commercial ProSim+ pour la simulation des procédés et de SimaPro® pour la partie simulation des impacts environnementaux. La matrice des résultats [impacts : procédés] est ensuite analysée par positionnement multidimensionnel en exploitant R (R Core Team, 2020). La Figure 4 montre l'enchaînement des outils, l'interopérabilité à couplage faible étant assurée par fichiers CSV.

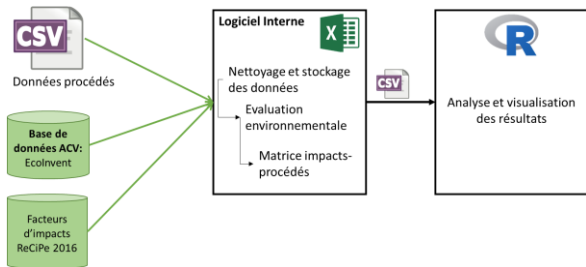


Figure 4: Schéma des outils logiciel

La dernière étape *Analyse et visualisation des résultats*, constituant le thème principal de ce papier, est maintenant décrite. Le positionnement multidimensionnel (MDS) classique, appelé aussi Principal Coordinates Analysis (PCoA) est une méthode de visualisation des (dé)similarités entre des objets dans un espace de dimension réduite. Elle est conçue pour permettre de comprendre les structures de proximités et d'opposition. Partant des informations sur les ressemblances mutuelles de n objets, souvent

avec la matrice de similarité ou $\Delta = \delta(i, j)_{0 \leq i \leq n; 0 \leq j \leq n}$, on cherche une configuration de n points (dans \mathbb{R}^2 en général) qui soit telle que si l'objet i ressemble à j plus que l'objet ℓ à k , on ait $\delta(i, j) < \delta(\ell, k)$. Pour n'importe

quelle matrice de distance de taille $n \times n$, la MDS permet de trouver un ensemble de n points repérés par leur coordonnées dont la matrice de similarité est égale ou très proche des données. Dans notre approche, nous utilisons la distance de Canberra (Lance and Williams, 1967, 1966), une version pondérée de la distance de Manhattan.

Soient x_r ($r = 1, \dots, n$) les coordonnées de n points dans un espace euclidien de p dimensions données où $x_r = (x_{r1}, x_{r2}, \dots, x_{rp})^T$ et $[B]_{rs} = b_{rs} = x_r^T x_s$. Pour une distance euclidienne $\Delta = \delta_{rs}$, à partir d'une matrice A d'éléments $[A]_{rs} = a_{rs} = -\frac{1}{2}\delta_{rs}^2$, déduite de la décomposition $x_r^T x_s$, la matrice B est obtenue en utilisant la relation suivante :

$$B = HAH \quad (1)$$

où H est matrice *centrage* : $H = I - n^{-1}\mathbb{1} \cdot \mathbb{1}^T$ avec $\mathbb{1} = (1, 1, \dots, n)^T$. Les éléments de A sont définis comme suite :

$$a_{r\bullet} = n^{-1} \sum_s a_{rs}$$

$$a_{\bullet s} = n^{-1} \sum_r a_{rs} \text{ et}$$

$$a_{\bullet\bullet} = n^{-1} \sum_r \sum_s a_{rs}$$

La procédure algorithmique du positionnement multidimensionnel peut être résumée, comme illustrée dans (Cox and Cox, 2001) par les étapes suivantes :

1. obtenir la matrice de proximité $\Delta = \delta_{rs}$
2. trouver la matrice $A = [-\frac{1}{2}\delta_{rs}^2]$
3. trouver la matrice $B = [a_{rs} - a_{r\bullet} - a_{\bullet s} + a_{\bullet\bullet}]$
4. trouver les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$ et les vecteurs propres v_1, v_2, \dots, v_{n-1}
si B est semi-définie positive (certaines valeurs propres sont négatives), soit (i) ignorer les valeurs

négatives et continuer, ou (ii) ajouter une constante appropriée c aux (dis)similarités

5. choisir un nombre de dimension approprié ℓ , utilisation possible de $\frac{\sum_1^\ell \lambda_i}{\sum(\text{valeurs propres positives})}$
6. les coordonnées des n points dans l'espace euclidien de dimension ℓ sont données par $x_{ri} = v_{ir}$ ($r = 1, \dots, n; i = 1, \dots, \ell$).

L'interprétation d'un résultat de MDS est le même que pour toute autre méthode de réduction de dimension : les objets qui sont plus proches sur le nuage de points sont plus semblables que ceux qui sont plus éloignés. C'est-à-dire que les points projetés sont disposés de telle manière que les points groupés (petite distance entre eux) reflètent les relations originales dans les données. Cependant, des informations supplémentaires sont nécessaires pour rendre la projection plus informative. Comme l'illustre la Figure 2, un algorithme de regroupement (k-moyennes) a été appliqué à la projection MDS pour mettre en évidence les objets les plus similaires (Impacts - Procédés). Les cartes des impacts et des procédés agrégés en deux dimensions sont présentées à la **Erreur ! Source du renvoi introuvable**. Figure 5. Pour le diagramme de dispersion des points d'impacts (Figure 5 gauche), nous pouvons distinguer trois groupes en utilisant des k-moyennes k :

- Groupe 1 : acidification terrestre (TA), eutrophisation de l'eau douce (FrEu), écotoxicité terrestre (TecoX) et écotoxicité de l'eau douce (FrEco). On retrouve dans ce groupe presque tous les impacts qui concernent la pollution chimique des sols et étendus d'eau.
- Groupe 2 : toxicité humaine (HT), formation de particules (PM), changement climatique partie santé humaine (CCHH), écotoxicité marine (MaEco), appauvrissement des métaux (MeDe), appauvrissement des

fossiles (FossDe), changement climatique partie écosystèmes (CCE) et rayonnement ionisant (IR). Ce groupe forme trois sous-groupes avec des points superposés (dans la perspective 2-D). Cela suggère que ces points sont très similaires en se basant sur la distance de Canberra. Ici, on retrouve un groupe assez hétérogène où se retrouvent les impacts non présents dans les groupes 1 et 3.

- Groupe 3 : occupation des terres urbaines (UrbLOcc), formation d'oxydants photochimiques (Pohto_ChOx), occupation des terres agricoles (AgLOcc), appauvrissement de la couche d'ozone (OD) et transformation naturelle des terres (NLTran). Dans ce groupe se retrouve principalement les impacts liés à l'occupation et la transformation des sols.

Des comparaisons similaires peuvent être faites entre tous les ensembles d'impacts. Les résultats de k-moyennes des éléments des procédés projetés (Figure 5 droite) sont similaires à ceux obtenus pour les impacts (Figure 5 gauche). Les acronymes légendant les points représentent chaque type de procédés. Ceux-ci ne seront pas explicités ici. Trois groupes distincts de procédés peuvent être identifier. Cependant, des clusters très serrés et très séparés apparaissent dans les données des procédés, ce qui peut suggérer que chaque cluster est un domaine ou un sous-domaine qui doit être analysé individuellement. Dans le groupe 3 par exemple on retrouve deux prétraitements purement mécanique (PM-UFM pour « pre-miling and ultra-fine milling »). En remontant jusqu'aux impacts on retrouve que ces deux prétraitements ont un impact sur l'appauvrissement des fossiles très important par rapport aux autres. Pour le groupe 1, les trois prétraitements ont des impacts relativement similaires pour tous les impacts et pour l'appauvrissement des fossiles on retrouve un impact autour de la dizaine de \$ alors que les prétraitements du groupe 2 ont un impact autour de 1\$. Alors que la visua-

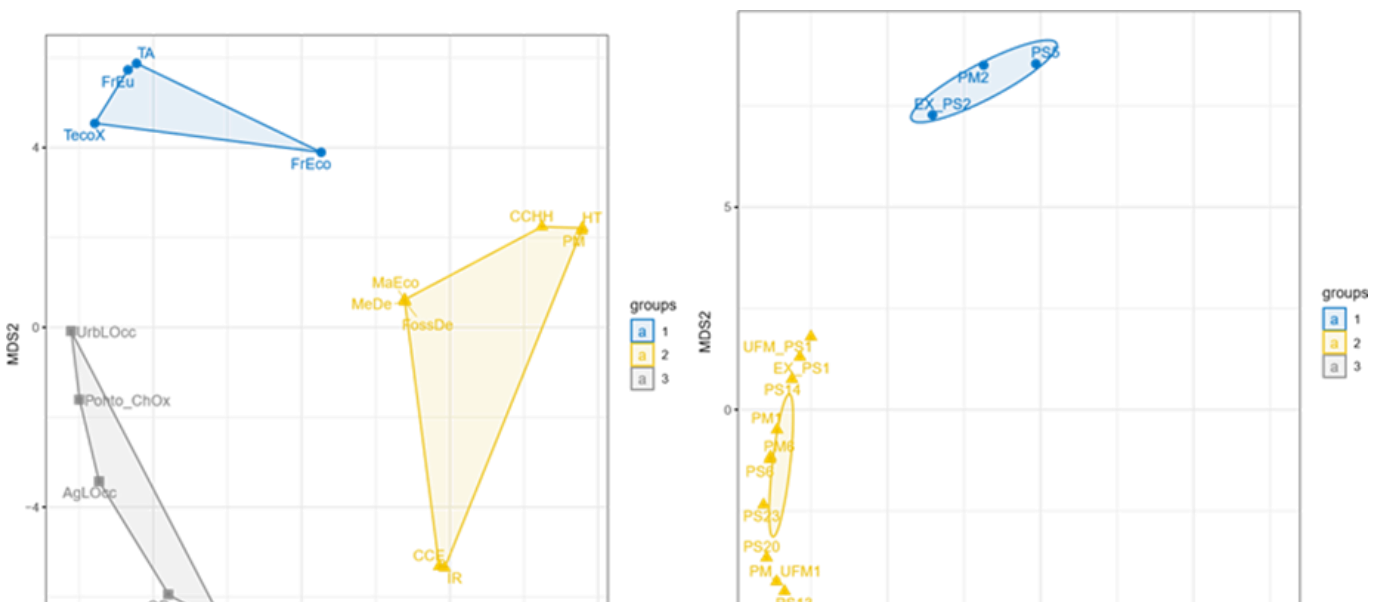


Figure 5 : Diagramme de dispersion de la projection (MDS) (deux dimensions) et regroupement des k-moyennes basé sur la matrice de distance des impacts (à gauche) et la matrice de distance des procédés (à droite).

lisation est le résultat de l'approche, on retrouve une interprétation que pourrait réaliser un expert ACV et procédés de prétraitement en exploitant la méthode ACV usuelle. Ces résultats ne sont cependant que partiels. En effet, on n'utilise qu'une représentation de la dimension 1 par rapport à la dimension 2 ce qui donne un pourcentage de représentativité faible et donc une perte d'informations. Il est envisagé de prolonger cette étude afin de représenter les dimensions 3 et 4 et compléter l'analyse.

4 CONCLUSION ET PERSPECTIVES

Une approche couplant la science des données et l'analyse environnementale est proposée. Composée de cinq étapes cette approche est présentée comme une aide pour le chercheur ou l'ingénieur R&D lors d'une étude préliminaire. Elle permet un gain de temps, une moindre expertise et aucune mise en œuvre expérimentale directe (partant du principe que les données scientifiques de premier plan existent et sont accessibles au sens « big data »). Cette approche est testée par l'étude de la valorisation de déchets agricoles pour promouvoir une économie circulaire en boucle locale. Cette application exécute les quatre premières étapes de l'approche. Elle compare différentes technologies de procédés de prétraitement de deux biomasses lignocellulosiques transformées en glucose, la tige de maïs et la paille de riz. Suite aux étapes de *but et frontière de l'étude, architecture des données et analyse environnementale*, une matrice impacts-procédés est analysée dans l'étape *visualisation et analyse des résultats* au travers de la méthode positionnement multidimensionnelle (MDS). L'approche fournit des premiers enseignements pour trier par groupes et établir une présélection de technologies ou de biomasses de manière rigoureuse et poursuivre le développement technologique sur des échelles de TRL supérieures.

Cette application révèle différentes limites. Les données issues de la littérature scientifique sont par nature des données issues de série d'expérimentations batch en laboratoire. L'analyse du cycle de vie est donc exécutée pour un niveau bas de maturité technologique (TRL 1/2). L'approche n'intègre pas le changement d'échelle nécessaire à la mise en œuvre d'un pilote semi-industriel, spécialement si le procédé devient semi-continu. Une autre limite est l'abondance et la qualité des données qui peuvent s'avérer insuffisantes pour des nouvelles voies technologues. D'autres axes de progrès sont de reconsidérer l'unité fonctionnelle, la stratégie de l'analyse environnementale globale par l'intégration de la phase amont agricole (ACV conséquente, politique d'allocation et d'extension de système) et la prise en compte de la chaîne logistique globale selon une analyse dynamique c'est-à-dire spatiale voire temporelle (Er-Raki et al., 2019). Particulièrement la biomasse lignocellulosique issue de l'amont agricole peut être d'une qualité et d'un approvisionnement variables.

Egalement, la politique des impacts relatifs aux effluents peut être modifiée. Tous les impacts sont attribués au produit final (le glucose). Or, selon le devenir de ces effluents

et leur valorisation au sens économie circulaire, l'interprétation des voies technologiques et des intrants peut différer.

REFERENCES

- Abdella, G.M., Kucukvar, M., Onat, N.C., Al-Yafay, H.M., Bulak, M.E., 2020. Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption. *J. Clean. Prod.* 251, 119661. <https://doi.org/10.1016/j.jclepro.2019.119661>
- Barry, C., 2020. La production de déchets non dangereux dans les industries agroalimentaires, Primeur. Agreste.
- Belaud, J.-P., Adoue, C., Vialle, C., Chorro, A., Sablayrolles, C., 2019a. A circular economy and industrial ecology toolbox for developing an eco-industrial park: perspectives from French policy. *Clean Technol. Environ. Policy* 21, 967–985. <https://doi.org/10.1007/s10098-019-01677-1>
- Belaud, J.-P., Prioux, N., Vialle, C., Sablayrolles, C., 2019b. Big data for agri-food 4.0: Application to sustainability management for by-products supply chain. *Comput. Ind.* 111, 41–50. <https://doi.org/10.1016/j.compind.2019.06.006>
- Bhinghe, R., Srinivasan, A., Robinson, S., Dornfeld, D., 2015. Data-intensive Life Cycle Assessment (DILCA) for Deteriorating Products. *Procedia CIRP* 29, 396–401. <https://doi.org/10.1016/j.procir.2015.02.192>
- Cooper, J., Noon, M., Jones, C., Kahn, E., Arbuckle, P., 2013. Big Data in Life Cycle Assessment: Big Data in Life Cycle Assessment. *J. Ind. Ecol.* 17, 796–799. <https://doi.org/10.1111/jiec.12069>
- Cox, T.F., Cox, M.A.A., 2001. Multidimensional scaling. Chapman & Hall/CRC, Boca Raton.
- Davis, C.B., Aid, G., Zhu, B., 2017. Secondary Resources in the Bio-Based Economy: A Computer Assisted Survey of Value Pathways in Academic Literature. *Waste Biomass Valorization* 8, 2229–2246. <https://doi.org/10.1007/s12649-017-9975-0>
- Ellen Macarthur Foundation, 2015. Towards a circular economy: business rationale for an accelerated transition. Ellen Macarthur Foundation.
- Er-Raki, A., Hartmann, D., Negny, S., Belaud, J.-P., 2019. Comparison of static and dynamic approaches in LCA: review and application. Presented at the International Conference on Industrial Ecology and Life Cycle Assessment, ICIELCA 2019, Rome, Italy.
- Frischknecht, R., Jungbluth, N., Althaus, H.-J., Doka, G., Dones, R., Heck, T., Hellweg, S., Hischier, R., Nemecek, T., Rebitzer, G., Spielmann, M., 2005. The ecoinvent Database: Overview and Methodological Framework (7 pp). *Int. J. Life Cycle Assess.* 10, 3–9. <https://doi.org/10.1065/lca2004.10.181.1>

- Grimaud, G., Perry, N., Laratte, B., 2017. Decision Support Methodology for Designing Sustainable Recycling Process Based on ETV Standards. *Procedia Manuf.* 7, 72–78. <https://doi.org/10.1016/j.promfg.2016.12.020>
- Kouchaki-Penchah, H., Nabavi-Pelesaraei, A., O'Dwyer, J., Sharifi, M., 2017. Environmental management of tea production using joint of life cycle assessment and data envelopment analysis approaches. *Environ. Prog. Sustain. Energy* 36, 1116–1122. <https://doi.org/10.1002/ep.12550>
- Lance, G.N., Williams, W.T., 1967. Mixed-data classificatory programs i - agglomerative systems. *Aust. Comput. J.* 1, 15–20.
- Lance, G.N., Williams, W.T., 1966. Computer programs for hierarchical polythetic classification (“similarity analyses”). *Comput. J.* 9, 60–64.
- Romeiko, X.X., Lee, E.K., Sorunmu, Y., Zhang, X., 2020. Spatially and Temporally Explicit Life Cycle Environmental Impacts of Soybean Production in the U.S. Midwest. *Environ. Sci. Technol.* 54, 4758–4768. <https://doi.org/10.1021/acs.est.9b06874>
- Zhao, L., Zha, Y., Zhuang, Y., Liang, L., 2019. Data envelopment analysis for sustainability evaluation in China: Tackling the economic, environmental, and social dimensions. *Eur. J. Oper. Res.* 275, 1083–1095. <https://doi.org/10.1016/j.ejor.2018.12.004>