

Concept d'empreintes chimiques appliqué à la gestion du risque chimique des matériaux, gisements recyclés et emballages alimentaires

*Concept of chemical fingerprints applied to chemical
risk management of materials, recycled deposits and
food packaging*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581 : agriculture, alimentation, biologie, environnement et santé
(ABIES)

Spécialité de doctorat : Chimie
Graduate School : Biosphera. Référent : AgroParisTech

Thèse préparée dans l'UMR SayFood (Université Paris-Saclay, INRAE, AgroParisTech), sous
la direction d'**Olivier VITRAC**, Directeur de Recherche, la co-direction de **Sandra
DOMENEK**, Professeure, le co-encadrement de **Cédric LYATHAUD**, Ingénieur de
Recherche, et de **Phuong Mai NGUYEN**, Ingénierie de Recherche (PhD)

Thèse soutenue à Paris-Saclay, le 21 mars 2023, par

Julien KERMORVANT

Composition du jury

Membres du jury avec voix délibérative

Pascal CARDINAEL Professeur, Université Rouen Normandie	Rapporteur & Examinateur
Thomas KARBOWIAK Professeur, Institut Agro Dijon	Rapporteur & Examinateur
Antoine BUREL Ingénieur de Recherche, ARKEMA	Examinateur
Valérie CAMEL Professeure, AgroParisTech (Université Paris-Saclay)	Examinaterice
Emmanuelle VULLIET Chargée de Recherche, CNRS (Université Claude Bernard Lyon 1)	Examinaterice
Florence CHURLAUD Ingénierie de Recherche, ARKEMA	Invitée

Titre : Concept d'empreintes chimiques appliqué à la gestion du risque chimique des matériaux, gisements recyclés et emballages alimentaire

Mots clés : Contact Alimentaire, Risque Chimique, Empreintes, Economie Circulaire, Recyclage

Résumé : L'économie circulaire remet en cause la possibilité de gérer collectivement les risques sanitaires des nouveaux matériaux et applications des emballages par les acteurs de la filière agroalimentaire. La contamination indirecte des aliments par les emballages est gérée par des concepts ambigus de substances intentionnellement ou non intentionnellement ajoutées, qui disparaissent quand les gisements sont mélangés ou vieillis. Les travaux visent à mesurer la complexité des mélanges de contaminants pour construire une empreinte de la contamination, d'un gisement de matériaux thermoplastiques ou cellulosiques. Afin de permettre une application en routine à l'échelle des acteurs de la filière (ex. mesure toutes les 20 tonnes, ou toutes les semaines), l'information chimique est quantifiée en chromatographie en phase gaz couplée à la

spectrométrie de masse basse résolution. Elle est enregistrée sous forme alphanumérique par une transformée en ondelettes qui code la variation du signal transformé, sa largeur et son intensité. Une substance ou un gisement peuvent être identifiés par une séquence préenregistrée dans une base de données. Une classification hiérarchique basée sur des critères complémentaires de distance d'édition et de concept d'information mutuelle permet l'évaluation du potentiel de la méthode pour tracer des gisements ou reconnaître des défauts sur ligne de production. Une application inverse permet de reconstruire un signal de chromatographie filtré pour être interprété avec les outils classiques de chimiométrie. Les résultats sont disponibles sous la forme d'un projet open source.

Title : Concept of chemical fingerprints applied to chemical risk management of materials, recycled deposits and food packaging

Keywords : Food Contact, Chemical Risk, Fingerprints, Circular Economy, Recycling

Abstract : The circular economy questions the possibility of collectively managing health risks caused by packaging new materials and applications by the actors of the food industry. The indirect contamination of food by packaging is managed through the ambiguous concepts of intentionally and non intentionally added substances, fading when deposits are blended or aged. The work aims at measuring the complexity of contaminants mixtures to build a fingerprint of the contamination, thermoplastic or cellulosic deposits. Chemical information is quantified with gas chromatography coupled with low-resolution mass spectrometry to enable a routine application at the scale of each

actor in the sector (ex. analysis every 20 t or each week). It is then registered as an alphanumerical chain by wavelet transform, encoding variations, widths, and intensities of the transformed signal. A preregistered sequence in a database could be used to identify a substance or a deposit. A hierarchical classification based on complementary edit distances and mutual information between samples enables the evaluation of the method's potential to trace deposits and pinpoint defaults on a manufacturing line. A reverse application allows the rebuilding of filtered chromatographic data for interpretation with classical chemometrics tools. All results are available as an open-source project.

Table des matières

1	Introduction	21
2	Bibliographie	23
2.1	Les matériaux au contact des aliments : une source systématique de contaminants dans les aliments	24
2.1.1	Quel terme choisir ? Contamination, migration, "outgassing ", une question de point de vue	24
2.1.2	Des règles disparates par matériau	25
2.1.3	Les plastiques : une utilisation et réutilisation très règlementées	27
2.1.4	Substances non intentionnellement ajoutées	29
2.1.4.1	Origine du terme	29
2.1.4.2	La contradiction environnementale	29
2.1.4.3	Une formalisation raisonnable	32
2.1.5	Calendrier des évolutions réglementaires	34
2.1.5.1	Calendrier français vs. calendrier européen	34
2.1.5.2	Les décrets 3R français : obligations et limites	36
2.2	Les matières recyclées au contact des aliments : une source additionnelle de contaminants	37
2.2.1	La pollution de la matière recyclée	37
2.2.2	Épuration des gisements : une alternative à la décontamination des gisements ?	39
2.2.3	Les limites de la décontamination des gisements	39
2.2.4	La gestion de la pollution pour la matière plastique recyclée	41
2.2.5	La barrière fonctionnelle : un concept pour maximiser la matière peu ou pas décontaminée	46
2.2.6	La difficulté du contrôle et de l'orientation des flux de matière recyclée	48
2.2.7	Ce que nous apprennent les crises passées sur les risques futurs : décalque, transferts sans contact	50
2.3	Apport des mathématiques à l'analyse des sources et mécanismes de contamination	54
2.3.1	La modélisation de la migration pour l'analyse des causes et des effets	54
2.3.2	La recherche de la signature chimique des contaminants	57
2.3.2.1	Le concept de source physique	57
2.3.2.2	Les techniques utilisées en chimiométrie	60
2.3.2.3	Les méthodes de poursuite	63
2.3.3	Ce que la bioinformatique pourrait apporter	64
2.3.4	Ce que la théorie de l'information pourrait apporter	66
2.4	Conclusions	67
3	Objectifs et démarche	71
3.1	Objectifs généraux	71
3.2	Objectifs spécifiques	74

3.3	Démarche	75
3.4	Théorie et choix du traitement du signal	76
3.4.1	Choix concernant les propriétés spécifiques aux signaux rencontrés	76
3.4.2	Choix sur la transformée en ondelettes continue	76
3.4.3	Choix sur l'encodage et la compression des données chromatographiques	81
4	Matériels et méthodes	83
4.1	Matériels	83
4.1.1	Échantillons de PP pour l'étude des effets de mise en forme et formulation	85
4.1.2	Échantillons de PP et PET issus de l'économie circulaire	86
4.1.2.1	Échantillons pour l'étude des origines de la matière recyclée	86
4.1.2.2	Échantillons pour l'étude de la traçabilité des gisements	86
4.2	Méthodes	87
4.2.1	Préparation et analyse chromatographique	87
4.2.1.1	Extraction, reconcentration et analyse en GC-MS	87
4.2.1.2	Thermodésorption et analyse en HS-GC-MS	89
4.2.2	Génération de chromatogrammes synthétiques	90
4.2.3	Codage de signaux chromatographiques	90
4.2.4	Encodage en séquences monotones	90
4.2.4.1	Justification	90
4.2.4.2	Règles de codage	91
4.2.5	Comparaison d'échantillons	91
4.2.5.1	Alignement des codes	91
4.2.5.2	Critère de Levenshtein	91
4.2.5.3	Critère de distance en entropie	92
5	Résultats et discussion	93
5.1	Modélisation du bruit associé aux signaux chromatographiques	94
5.1.1	Temporalisation des signaux chromatographiques	94
5.1.2	Description d'un chromatogramme intégré typique	96
5.1.3	Description du chromatogramme en deux dimensions correspondant	98
5.1.3.1	Modèles du détecteur et du signal mesuré $I_{t,m}$	100
5.1.3.2	Interprétation statistique des événements liés à l'échantillon $s_m(t)$	102
5.1.3.3	Détectabilité d'un ion issu d'une substance à faible concentration	105
5.1.3.4	Discussion sur la validité de l'hypothèse de Poisson pour le signal colonne $b_m^{col}(t)$	109
5.1.3.5	Conclusions de l'analyse statistique du signal chromatographique brut	111
5.2	Transformée de Ricker et encodage des signaux chromatographiques	112
5.2.1	Transformée de Ricker	112
5.2.1.1	Apport attendu de la transformation du signal	112
5.2.1.2	Signaux typiques transformés	113
5.2.1.3	Explication de l'effet de loupe de la transformée sur le signal	115
5.2.1.4	Analyse statistique du signal transformé	118

5.2.2	Encodage du signal transformé	121
5.2.2.1	Principe et propriétés de l'encodage en segments monotones	121
5.2.2.2	Statistiques de l'encodage et critère de rejet du bruit	123
5.2.2.3	Discussion de l'unicité de l'encodage entre répétitions d'un échantillon	125
5.2.2.4	Discussion de la différence d'encodage entre deux échantillons	127
5.2.2.5	Discussion sur les performances de la comparaison en caractères répétés et compressés	128
5.2.2.6	Résumé de l'encodage et de la comparaison	129
5.2.3	Propriétés et limitations des concepts mathématiques appliqués	131
5.3	Application aux polyoléfines vierges et recyclées	132
5.3.1	Vue générale	132
5.3.2	Identification des effets contrôlant la qualité de pots et de couvercles en PP	135
5.3.2.1	Plan d'échantillonnage	135
5.3.2.2	Filtrage et orthogonalisation de l'information portée par les combinaisons étudiées	138
5.3.2.3	Analyse des effets sur l'information chimique extraite par voie d'analyse par solvant	145
5.3.2.4	Analyse des effets sur l'information chimique extraite par voie d'analyse en phase gazeuse	149
5.3.3	Une première tentative de validation et de généralisation : comparaison de l'information chimique extraite d'échantillons de PP recyclés	154
5.3.3.1	Passage d'une cartographie locale (sur site de production) à une cartographie globale dans le cadre de l'économie circulaire	154
5.3.3.2	Calibration de la cartographie des PP recyclés	156
5.3.3.3	Reconstruction de la traçabilité et cartographie des PP recyclés	158
5.3.3.4	Ce qu'il faut retenir de la comparaison des PP et leurs gisements de matières recyclées	163
5.4	Application aux polyesters vierges, recyclés et décontaminés	165
5.4.1	Vue générale	165
5.4.2	Calibration de la cartographie du marché circulaire mondial des PET	168
5.4.3	Reconstruction de la traçabilité et cartographie du marché circulaire mondial des PET	170
5.4.4	Ce qu'il faut retenir de la comparaison des gisements de PET recyclés et de leur décontamination	175
5.4.4.1	Ce qui change quand la cartographie chimique est étendue à toute une filière	175
5.4.4.2	Espace chimique et flux d'information dans une boucle de recyclage	176
5.5	Discussion Générale	179
5.5.1	Préambule	179
5.5.1.1	Des valeurs de γ qui sont largement au-delà de nos estimations initiales : des risques émergents probables au-delà de la problématique initiale des NIAS	179

5.5.1.2	∇ : un concept possible pour gérer l'urgence réglementaire	180
5.5.1.3	Éléments de la discussion	181
6	Conclusion et perspectives	187
6.1	Conclusions générales	187
6.1.1	La reformulation du problème d'identification et de gestion des NIAS	187
6.1.2	L'introduction de notions thermodynamiques	188
6.1.3	Le codage du signal chromatographique et son interprétation	189
6.1.4	La comparaison de codes « empreintes » entre échantillons	190
6.1.5	Le concept d'espace chimique et de déplacement chimique dans cet espace	191
6.1.6	L'espace chimique dans le cadre de l'économie circulaire	192
6.2	Perspectives	193
6.2.1	Le nombre de l'économie circulaire : ∇	193
6.2.2	L'hypothèse de détectabilité des transferts via l'échange des informations	193
6.2.3	Implémentation et utilisation de la méthode dans les plans de surveillance de la matière recyclée	194
6.2.4	Le partage de l'information chimique et la coopération des acteurs	194

Liste des figures

2.1	Hétérogénéités de principes et de réglementations en Europe pour les différents composants de l'emballage	26
2.2	Des matériaux plastiques au contact des aliments aux déchets	28
2.3	Controverses associées à l'utilisation de la matière recyclée pour le contact alimentaire	30
2.4	Comparaison des protocoles de gestion de la sécurité sanitaire des aliments et de leur emballage	33
2.5	Calendrier des échéances imposées en France et en Europe	35
2.6	Vue générale des niveaux de contamination des aliments par les substances issues des emballages en papier et carton	38
2.7	Mécanismes et cascade chimique entraînant une contamination variable en espèces et quantités	40
2.8	Intérêt de l'utilisation d'une barrière fonctionnelle dans le design d'un emballage incorporant de la matière recyclée	47
2.9	Complexité du contrôle et de l'orientation des flux illustrés par des exemples	48
2.10	Exemple de transfert de matière entre matériaux et de contamination de l'aliment sans contact	53
2.11	Principes reconnus et méthodologie pour démontrer la conformité des matériaux au contact des aliments via une modélisation itérative	56
2.12	Classification des échantillons d'après les trois méthodes utilisées par Kassouf et al. (2013)	62
3.1	Ondelettes hermitiennes d'ordres 1, 2 et 3	78
3.2	Application de la transformée en ondelettes avec les ondelettes hermitiennes d'ordres 1, 2 et 3 sur un ersatz de signal chromatographique	79
3.3	Influence de l'échelle sur la transformée	79
3.4	Échelle optimale en fonction de la largeur d'une gaussienne	80
3.5	Encodage des six segments possibles et d'un pic	81
3.6	Compression obtenue par l'encodage du signal transformé en fonction de l'échelle	82
4.1	Échantillons pour l'étude des origines de la matière recyclée	86
5.1	Acquisition d'un chromatogramme par un appareil GC-MS (Quadrupôle)	95
5.2	Signal TIC typique d'un extrait de PET recyclé en GC-MS (13330 points) et programme de température du four	96
5.3	Principe de la conservation du signal TIC de la Figure 5.2 en indice de rétention.	97
5.4	Représentation en fausses couleurs du signal correspondant en 2D	98
5.5	Isocontours de l'inégalité de Chernoff entre un signal d'intérêt $s_m(t)$ et une série d'événements poissonniens $b_m^{col}(t)$	103
5.6	Exemples de trois comportements de ligne de base	105

5.7	Signal 2D correspondant filtré avec le critère de Bienaymé-Tchebychev avec intervalle de confiance unilatéral à 5%	107
5.8	Signal 2D correspondant filtré avec le critère de Chernoff avec intervalle de confiance unilatéral à 50%	108
5.9	Caractérisation des ions selon leur ligne de base et leur position	109
5.10	Signaux brut et transformée en ondelette aux échelles 1,8 et 64 des ions 83, 114 et 319	114
5.11	Amplification maximale théorique en fonction de la largeur des signaux gaussiens et de l'optimalité de l'échelle de la transformée et comparaison avec signal réel	116
5.12	Transformée en ondelette aux échelles 1,8 et 64 des ions 83, 114 et 319 avec enveloppe du critère de Bienaymé-Tchebychev	119
5.13	Comparaison de l'encodage à l'échelle $\ell = 8$ des répétitions de l'échantillon de PET vierge G5-1	126
5.14	Comparaison de l'encodage à l'échelle $\ell = 8$ de du PET vierge G5-1 avec le PET recyclé non décontaminé G15-1	127
5.15	Comparaison des TIC d'un vPP et d'un rPP en GC-MS	133
5.16	Classification des échantillons de polypropylène mis en forme basée sur les effets <i>a priori</i>	136
5.17	MSD du cumul des dimensions pour les articles en PP	139
5.18	Comparaison des distances inter échantillons avec la distance inter répétitions	140
5.19	Premier centile significatif par dimension	142
5.20	ANOVA des effets escomptés pour le suivi d'articles en PP	143
5.21	Exemples de projections dans l'espace généré par la distance de Levenshtein	146
5.22	Exemples de projections dans l'espace généré par la distance en entropie	147
5.23	MSD associée au cumul des dimensions en pourcentage de la MSD totale pour les distances de Levenshtein et entropique en analyse HS-GC-MS.	149
5.24	Premier centile significatif par dimension en HS-GC-MS	150
5.25	ANOVA des effets escomptés pour le suivi de ligne de production en HS-GC-MS	151
5.26	Exemple de projection dans l'espace généré par la distance de Levenshtein en HS-GC-MS	152
5.27	Exemple de projection dans l'espace généré par la distance en entropie en HS-GC-MS	152
5.28	Classification des échantillons de polypropylène vierges et recyclés basée sur les effets <i>a priori</i>	155
5.29	MSD du cumul des dimensions pour les polypropylènes vierges et recyclés	156
5.30	Premiers centiles significatifs pour l'espace des PP recyclés	157
5.31	ANOVA à trois effets (imbriqués et isolés) sur les PP recyclés	158
5.32	Cartes des PP recyclés en distance de Levenshtein	160
5.33	Cartes des PP recyclés en distance entropique	161
5.34	Classification <i>a posteriori</i> des PP recyclés selon la distance de Levenshtein	162
5.35	Comparaison des TIC d'un vPET et d'un rPET en GC-MS	166
5.36	Classification <i>a priori</i> des PET recyclés	167
5.37	MSD du cumul de dimension appliquée à l'économie circulaire du PET	168
5.38	Premier quantile significatif par dimension pour l'économie circulaire du PET	169
5.39	Premier quantile significatif par dimension pour l'économie circulaire du PET	170
5.40	Projection des PET sur les dimensions 1×2 selon le critère de Levenshtein	171

5.41	Projection des PET sur les dimensions 1×2 selon le critère entropique	172
5.42	Classification <i>a posteriori</i> des PET recyclés selon la distance de Levenshtein	172
5.43	Classification <i>a posteriori</i> des PET vierges et recyclés selon la distance entropique . .	173
5.44	Exemple de chemin Γ de recyclage de PET	176
5.45	Cartographie des emballages et aliments selon le critère de Levenshtein	182
5.46	Modèle de l'aire de la transformée à l'échelle optimale en fonction de l'aire de la gaussienne initiale	184

Liste des tableaux

2.1	Valeurs seuils de concentration basées sur l'approche TTC pour un adulte de 60 kg et un apport journalier de 1 kg	41
2.2	Profile des contaminants issus de PET en contact avec des produits non-alimentaires	43
2.3	Mécanismes de transferts croisés entre matériaux capables de redistribuer les contaminants dans la filière de recyclage.	51
2.4	Résumé des principaux concepts de la modélisation de la migration	55
2.5	Cas d'étude types pour le concept de sources	57
2.6	Techniques communes utilisées en chimiométrie	60
2.7	Composés détectés selon les techniques chimiométriques utilisées par Kassouf et al. (2013)	62
2.8	Exemple d'alignement des mots niche et chiens issu de Wikipedia	65
2.9	Tableau récapitulatif de types de contaminants et leurs risques sanitaires associés	69
3.1	Démarche mise en place pour répondre aux objectifs	75
3.2	Propriétés des ondelettes hermitiennes d'ordres 1, 2 et 3 hermitiennes	77
4.1	Vue générale des échantillons utilisés pour construire les cartes des déplacements chimiques	84
4.2	Echantillons pour le suivi de procédé	85
4.3	Échantillons pour l'étude de la traçabilité des gisements	86
4.4	Paramètres de l'évaporateur	87
4.5	Solution d'étalons internes dans du chloroforme deutéré pour analyse en GC-MS	87
4.6	Séquence type en GC-MS	88
4.7	Solution d'étalons internes dans du méthanol pour analyse en HS-GC-MS	89
4.8	Séquence type en HS-GC-MS	89
4.9	Table d'encodage des segments monotones	91
5.1	Application des critères de Bienaymé-Tchebychev et Chernoff pour la détection de zones d'intérêt	107
5.2	Variances du signal colonne $b_{t,m}^{col}$, à basses et hautes températures, brut et avec transformée en ondelette aux échelles 1, 8 et 64	116
5.3	Comparaison de l'application du critère de Bienaymé-Tchebychev avant et après transformée de Ricker	120
5.4	Proportions des segments monotones aux échelles 1, 8 et 64	123
5.5	Proportions des segments monotones aux échelles 1, 8 et 64 après filtrage des segments non signifiant	124
5.6	Comparaison de la discernabilité des méthodes répétées et compressées	128
5.7	Dimensions d'intérêt pour les articles en PP analysés en GC-MS	141
5.8	Dimensions d'intérêt pour expliquer les effets escomptés du suivi d'articles en PP	145

5.9 Dimensions d'intérêt pour expliquer les effets escomptés du suivi de production en HS-GC-MS	150
5.10 Échantillons utilisés pour l'étude préliminaire de la contamination entre aliments et emballage	181

Liste des propriétés et propositions

3.1	Propriété (Linéarité de la transformée en ondelettes)	76
3.2	Propriété (Propriété de loupe continue sur signal transformé)	77
3.3	Propriété (Distribution de l'opérateur de dérivation dans la transformé)	77
4.1	Propriété (Caractéristiques d'un segment monotone)	91
5.1	Propriété (Spécificités de la transformée de Ricker)	118
5.1	Proposition (Critère de sélection des segments d'intérêt)	123

Tables des abréviations

Abréviation	Définition
ABA	Tricouche dont le matériau interne (B) est différent des faces externes (A)
ADEME	Agence de l'environnement et de la maîtrise de l'énergie
ADN	Acide désoxyribonucléique
IAEA	Agence internationale de l'énergie atomique (international atomic energy agency)
AGEC	Anti-gaspillage et économie circulaire (loi)
AMI	Appel à manifestation d'intérêt
ANOVA	Analyse de la variance (analysis of variance)
ANR	Agence nationale de la recherche
ARN	Acide ribonucléique
BfR	Institut fédéral allemand d'évaluation des risques (Bundesinstitut für Risikobewertung)
BT	Bienaymé-Tchébychev (critère de)
C7C40	Mélange d' alcanes linéaires allant de C7 (heptane) au C40 (tétracontane) par incrément de un
CASRN	Numéro CAS (chemical abstract service registry number)
CCSWA	Analyse en composantes communes et poids spécifiques (common components and specific weights analysis)
CE	Commission européenne
CMR	Cancérogène, mutagène et reprotoxique
DGCCRF	Direction générale de la concurrence, de la consommation et de la répression des fraudes
DIPN	Diisopropylnaphtalène (isomères)
EFSA	Autorité européenne de sécurité des aliments (european food safety authority)
EI	Étalon interne
ESA	Agence spatiale européenne (european space agency)
FDA	Agence fédérale américaine des produits alimentaires et médicamenteux (food and drug administration)
FMEA	Analyse des modes de défaillance, de leurs effets et de leur criticité (failure mode and effect critical analysis)
GC-MS	Chromatographie en phase gazeuse-spectrométrie de masse (gas chromatography-mass spectrometry)
HACCP	Système d'analyse des risques et de maîtrise des points critiques (hazard analysis critical control point)
HS	Espace de tête (headspace)
IARC	Centre international de recherche sur le cancer (international agency for research on cancer)
IAS	Substance intentionnellement ajoutée (intentionnally added substance)
IC	Chromatogramme d'un ion (ion chromatogram)
ICA	Analyse en composantes in dépendantes (independant components analysis)
IML	Étiquetage dans le moule (in-mould-labelling)
ITX	Isopropylthioxanthone

Abréviation	Définition
LATAM	Amérique latine (Latin America)
LNE	Laboratoire national de métrologie et d'essais
LOD	Limite de détection (limit of detection)
LOQ	Limite de quantification (limit of quantification)
MSD	Carré de la distance à la moyenne (mean square displacement)
MZ	Rapport de masse sur charge
NASA	Administration nationale de l'aéronautique et de l'espace (national aeronautics and space administration)
ND	Non détecté (non detected)
NIAS	Substance non intentionnellement ajoutée (non intentionally added substance)
NORAM	Amérique du Nord (North America)
NQ	Non quantifié (non quantified)
OCDE	Organisation de coopération et de développement économiques
ONU	Organisation des nations unies
PCA	Analyse en composantes principales (principal components analysis)
PCB	Polychlorobiphényles
PCoA	Analyse en coordonnées principales (principal coordinates analysis)
PET	Polyéthylène téréphthalate
PP	Polypropylène
PVC	Polychlorure de vinyle (polyvinyle chloride)
RI	Indice de rétention (retention index)
RMN	Résonance magnétique nucléaire
RT	Temps de rétention (retention time)
rX	Matériau X recyclé
TIC	Chromatogramme total (total ion chromatogram)
TTC	Seuil de préoccupation toxicologique (threshold of toxicological concern)
UE	Union européenne
UV	Ultra-violet
vX	Matériau X vierge (non recyclé)

Remerciements

Je tiens à exprimer toute ma reconnaissance aux personnes sans qui ces travaux n'auraient pas été possibles :

Tout d'abord à l'équipe de direction et d'encadrement de ma thèse, mon directeur Olivier Vitrac, ma codirectrice Sandra Domenek, mes encadrants Cédric Lyathaud et Phuong Mai Nguyen, pour leurs précieux conseils, leur soutien technique et moral ainsi que pour le temps passé sur ce manuscrit.

Aux membres du jury, Pascal Cardinael, Thomas Karbowiak, Antoine Burel, Valérie Camel, Emmanuel Vulliet et Florence Churlaud, qui ont accepté de prendre de leur temps pour s'intéresser à mes travaux et les évaluer.

A l'ensemble des équipes et collègues du LNE, d'AgroParisTech et de l'INRAE pour leur accueil et leur bienveillance. Ainsi qu'aux UMT SAFEMAT 17.09 puis 22.07 qui ont permis une dynamique et une collaboration dans lesquels mes travaux ont été intégrés.

Aux industriels et entités des domaines de la plasturgie, emballages et recyclage qui nous ont fait confiance et fournissent les ressources et échantillons nécessaires à la réalisation de ces travaux.

A mes amis pour les bons moments passés ensemble et leur support infaillible.

Et enfin à ma famille pour leurs encouragements et leur soutien permanent.

Préambule

Publié

Ces travaux ont fait l'objet d'une publication intitulée "Coding and tracking contaminants of food contact materials by wavelet analysis" dans les proceedings de la conférence FOODSIM'2022 qui s'est tenue à Gant, Belgique du 3 au 6 Avril 2022. Celle-ci est disponible en libre accès sur l'archive ouverte HAL à cette adresse : <https://hal.science/hal-03759593>.

L'ensemble des codes développés sur MATLAB sont disponibles en accès libre sous la licence publique générale GNU v3.0 sur le dépôt GitHub Circular_Daleth à cette adresse : https://github.com/jkermorv/Circular_Daleth. Certaines fonctions peuvent cependant nécessiter des ToolBoxes de MathWorks.

Soumis

Une seconde publication intitulée "L'économie circulaire et la sécurité sanitaire des emballages alimentaires : pourquoi elles s'opposent, pourquoi elles doivent aller de concert ?" à été soumise fin 2022 dans les Cahiers de Nutrition et de Diététique.

En préparation

Une publication sur les capacités de la méthode développée à comparer des chromatogrammes est en cours de rédaction. Une autre est prévue sur l'application pour une veille sanitaire dans le cadre de l'économie circulaire.

1 - Introduction

Les emballages alimentaires sont devenus omniprésents au fur et à mesure du détachement entre la production et la consommation alimentaire. En effet le mode de vie des pays développés pousse à éloigner les lieux de production des zones de consommation, rendant obligatoire le transport de denrées où l'emballage joue un rôle primordial pour le lotage et la protection des aliments. Cette évolution bienvenue introduit cependant un risque nouveau de contamination pour les consommateurs. Ce type d'exposition remonte au moins à l'Empire romain, période au cours de laquelle les demeures des plus fortunés étaient alimentées en eau par de la tuyauterie en plomb responsable d'une intoxication chronique menant au saturnisme, un exemple plus récent concerne les boîtes de conserve dont les soudures étaient elles aussi à base de plomb. Seules les contaminations menant à des intoxications aiguës étaient potentiellement détectables à l'époque, il a fallut attendre le développement de la chimie analytique et de la toxicologie pour espérer identifier les causes de maladies chroniques. La chimie analytique se limitait à ses début à l'analyse élémentaire utilisée en particulier pour la quantification des métaux lourds responsables de maladies chroniques, l'analyse de composés organique devra attendre le XX^{ème} siècle et plus particulièrement les années 70 avec le couplage des techniques séparatives avec des détecteurs.

De nos jours les emballages alimentaires sont plus variés tout comme les contaminations résultantes dont voici quelques exemples bien identifiés et réglementés : phtalates dans les emballages en plastique, huiles minérales dans les papiers et les cartons, agents gonflants dans les caoutchoucs. Cependant aucun risque non identifié n'est réglementé aujourd'hui. Le terme de danger caractérise les conséquences sur un être vivant que peut avoir l'exposition à une substance donnée, le danger peut être aiguë ou chronique, dans ce dernier cas il devient difficile de le mettre en évidence. Le risque quant à lui désigne l'occurrence d'une substance, sa concentration ou l'exposition d'une population. En Europe, la gestion du risque se base sur l'occurrence d'une substance dans l'aliment et sur une exposition chronique surestimée par consommation de 1 kg d'aliment au contact avec 6 dm² d'emballage. Aux États-Unis, l'exposition est estimée pour être représentative du marché actuel. Nous assistons aujourd'hui à l'émergence de nouveaux risques auparavant inconnus : perturbateurs endocriniens, effets liés à de faibles doses (la réponse toxicologique d'une substances n'est pas monotone avec la concentration), effets cocktails qui peuvent augmenter ou réduire la toxicité d'un mélange de substances par rapport aux substances séparées. De plus avec l'émergence de l'économie circulaire liée à la prise de conscience environnementale et à l'évolution de la réglementation de nouveaux enjeux sont à prendre en compte pour la gestion du risque chimique. Les boucles de recyclages pour les emballages alimentaires entraînent une incertitude sur l'origine des gisements et leur qualité nécessitant des précautions supplémentaires. Pour les plastiques, la décontamination est obligatoire et rend l'exposition du consommateur acceptable mais ne permet pas de faire disparaître totalement un contaminant du gisement. Pour les papiers et cartons, la décontamination n'est pas obligatoire, mais un désencrage peut être envisagé, cette politique a entraîné la contamination de l'ensemble de la boucle de recyclage par des composés comme les isomères du diisopropylnaphtalène (DIPN).

Nous pensons qu'il est nécessaire de suivre cette évolution du marché en rendant contrôlables les pratiques industrielles à petites et grandes échelles, de la ligne de production à la boucle de recyclage complète. Nous nous retrouvons aujourd'hui dans un cas similaire à la génération spontanée observée à la fin du XIX^{ème} siècle en biologie, des contaminations inévitables, avec des causes mal définies ou incomprises et un bestiaire des possibilités non exhaustif. Nos connaissances et nos capacités analytiques ne nous permettent de détecter et d'identifier qu'un nombre limité de composés dans les migrants, responsables que d'une infime partie de la réponse toxicologique totale. La prédominance de certaines disciplines telles que la chimie analytique et de la toxicologie au détriment ou sans relation avec la science des polymères et des transferts limite grandement notre compréhension de la problématique dans son ensemble. La variabilité des substances non intentionnellement ajoutées (NIAS) tant dans leurs formes que leurs proportions est telle que nous pensons nécessaire de rompre les stratégies mises en place aujourd'hui pour se concentrer sur une approche globale basée sur des empreintes pouvant muter ou être transférées à l'instar du matériel génétique.

Ce manuscrit débutera par une bibliographie pour introduire et discuter des concepts de NIAS, de théorie de l'information ainsi que des pratiques de traitement du signal et de chimiométrie. Dans un deuxième temps, les objectifs ciblés, les démarches mises en place et les résultats attendus seront précisés. Les méthodologies suivies et les cas d'études seront ensuite présentés. Ils seront suivis d'une présentation et d'une discussion autour des principaux résultats. Enfin les propriétés et limitations des méthodes seront résumées, avant de proposer des perspectives aux travaux réalisés.

2 - Bibliographie

Le contexte d'économie circulaire a profondément modifié le concept de contamination et de substance non intentionnellement ajoutée (NIAS) y compris parmi pendant la durée des travaux. Nous sommes passés d'une notion de NIAS fortuit à une notion de contamination systématique par des substances inconnues dont la thermodynamique prédit une augmentation rapide du nombre de substances présentes dans le matériau et susceptibles d'être transférées. Ce chapitre fait le point sur le glissement des concepts et la crainte que suscite la généralisation de la matière recyclée au contact des aliments, qu'elle soit d'origine pétrochimique ou cellulosique. La première partie s'intéresse aux racines des préoccupations des autorités et des metteurs sur le marché : les mécanismes de transferts du matériau vers l'aliment et de leurs contrôles. Nous parlons généralement d'inertie, mais ce concept se décline en quantité et type de substances. Les deux aspects sont affectés par les nouvelles obligations de circularité. Est-il raisonnable d'imaginer qu'une substance fortuite passe en dessous des radars ? Est-ce que les traces de contamination peuvent être effacées ? Est-ce que le profil des contaminants ne va pas augmenter avec l'arrivée de nouvelles pratiques et de nouveaux matériaux ?

Des éléments de réponse à ces questions seront présentés dans la Partie 2.2. Les Parties 2.1 et 2.2 s'appuient sur la publication (Kermorvant et al., 2023), qui décrit avec plus de détails les conséquences attendues sur la circularité des matériaux au contact des aliments. Ces risques émergents mal établis peuvent être perçus comme angoissants. Le rôle du scientifique est de les rationaliser. Les mathématiques ont été proposées très tôt pour gérer les risques de contamination non pas parce qu'elles étaient nécessairement précises et justes à prévoir les risques, mais parce qu'elles offraient un cadre pour extrapoler des effets (temps, température, géométrie, concentration, aliment) sur la valeur de la contamination de l'aliment et finalement de l'exposition du consommateur. Cette objectivité intrinsèque des mathématiques est discutée dans la troisième partie pour le problème dit de NIAS. Comment comparer des matériaux proches, mais présenter des différences que l'on ne sait pas décrypter complément ?

2.1 . Les matériaux au contact des aliments : une source systématique de contaminants dans les aliments

2.1.1 . Quel terme choisir ? Contamination, migration, "outgassing ", une question de point de vue

Les emballages plastiques tout comme la grande majorité des matériaux au contact des aliments ne sont pas inertes. Ils transfèrent des substances aux aliments tout comme tous les matériaux qui rentrent dans un véhicule spatial relarguent des substances qui risquent d'opacifier les optiques. Toutes substances non liées de manière covalente avec les chaînes de polymère sont potentiellement transférables et pourront réagir avec les autres constituants des aliments et des emballages. La NASA ([NASA, 2018](#)) et l'ESA ([ESA, 2018](#)) maintiennent des bases de données sur les capacités de relargage "outgassing" de dizaines de milliers de matériaux de tous types (plastiques, colles, adhésifs...). Les mécanismes ne sont pas fondamentalement différentes les substances peuvent être transférées même sans contact si elles ont suffisamment de vitesse ([Nguyen et al., 2019](#)) ou au travers d'autres matériaux, si on leur laisse suffisamment de temps ([Nguyen et al., 2013](#)).

La volatilité de la substance accroît le transfert uniquement sur les temps courts, car sur les temps longs les équilibres thermodynamiques sont déplacés par la condensation des substances sur une paroi froide ou par les constituants de l'aliment ([Nguyen et al., 2017](#)). Le terme "migration" a rendu la contamination des aliments par les matériaux plus acceptable. Les phénomènes de "migration" et d'interactions contenant-contenu ont été reconnus très tôt par les autorités tout comme le danger réel que pouvait poser l'ingestion des substances issues des matières plastiques a été identifié dès le début de l'histoire des emballages alimentaires. L'emballage se développe d'abord sur le modèle de consommation américain, très loin des préoccupations environnementales actuelles. Les formes des bouteilles plastiques et des sachets ont peu varié pendant des décennies. La chimie des matériaux va-t-elle être modifiée en profondeur sous l'effet de l'action réglementaire combinée des pouvoirs publics, des études scientifiques et plus tardivement de la prise de conscience des consommateurs. Aujourd'hui, les bouteilles ne sont plus en polychlorure de vinyle (PVC), mais en polyéthylène téraphthalate (PET), le styrène et l'éthylbenzène récemment classées comme potentiellement cancérogènes par l'Agence des Nations Unies (IARC) ([IARC, 2019](#)) accélèrent la substitution des matériaux des pots de yaourt en France. Les phtalates ne sont plus un additif ajouté intentionnellement, mais un contaminant exogène des plastiques ou une impureté issue des boucles de recyclages.

2.1.2 . Des règles disparates par matériau

L'existence d'un danger est reconnue par les autorités européennes dès 1976 avec les principes généraux énoncés dans la directive 76/893/CEE puis repris et étendus dans les règlements cadres successifs (directive 89/109/CEE, règlements (CE) n° 1935/2004 et (CE) n° 2023/2006) : Sont concernés :

- Les matériaux et objets qui, à l'état de produits finis, sont destinés à être mis en contact ou sont mis en contact, conformément à leur destination, avec des denrées alimentaires.
- Les matériaux et objets en contact avec l'eau destinée à la consommation humaine (hors systèmes d'adduction d'eau).
- Les matériaux et objets doivent être fabriqués conformément aux bonnes pratiques de fabrication afin que, dans les conditions normales ou prévisibles de leur emploi, ils ne cèdent pas aux denrées alimentaires des constituants dans une quantité susceptible : de présenter un danger pour la santé humaine, d'entrainer une modification inacceptable de la composition des denrées alimentaires ou une altération des caractères organoleptiques de celles-ci.

Des dispositions particulières [sont] applicables à certains groupes de matériaux et objets. Ces directives [ou règlements] spécifiques peuvent comporter notamment :

- Si possible et si nécessaire, la liste des substances et matières dont l'emploi est autorisé à l'exclusion de toutes autres [liste positive] ;
- les critères de pureté de ces substances et matières ;
- les conditions particulières d'emploi de ces substances et matières et/ou des matériaux et objets dans lesquels ces substances et matières ont été utilisées ;
- des limites spécifiques de migration de certains constituants ou groupe de constituants dans ou sur les denrées alimentaires ;
- une limite globale de migration des constituants dans ou sur les denrées alimentaires.

Bien que la sécurité sanitaire soit reconnue à l'échelle européenne, l'arbre réglementaire mis en place ne dépend pas directement de la substance, du niveau de consommation ou du consommateur ciblé (enfant, adulte), mais du matériau à l'origine de la substance. Cette construction résulte d'un glissement vers d'un laissez-faire vers des règles spécifiques sous la pression des crises passées, des consommateurs et des priorités nationales. La construction protège d'abord la liberté de circulation au sein de l'Union Européenne des composants des emballages qu'ils soient issus de la chimie de synthèse (polymères, additifs, solvants, auxiliaires technologiques, encres, colles et adhésifs, vernis), de la biomasse (matériaux cellulosiques, bois, caoutchouc) ou des minéraux (verre). La Figure 2.1 montre la différence d'appréciation pour les différents matériaux entre danger, occurrence et niveau de réglementation. La conséquence pratique est que des principes différents prévalent pour les composants d'un emballage composite. La protection est plus forte pour les matières plastiques que pour les autres matériaux avec des obligations d'autorisation de substances et d'exposition faible aux substances les plus toxiques via des limites de migration qui doivent être vérifiées sur le produit fini.

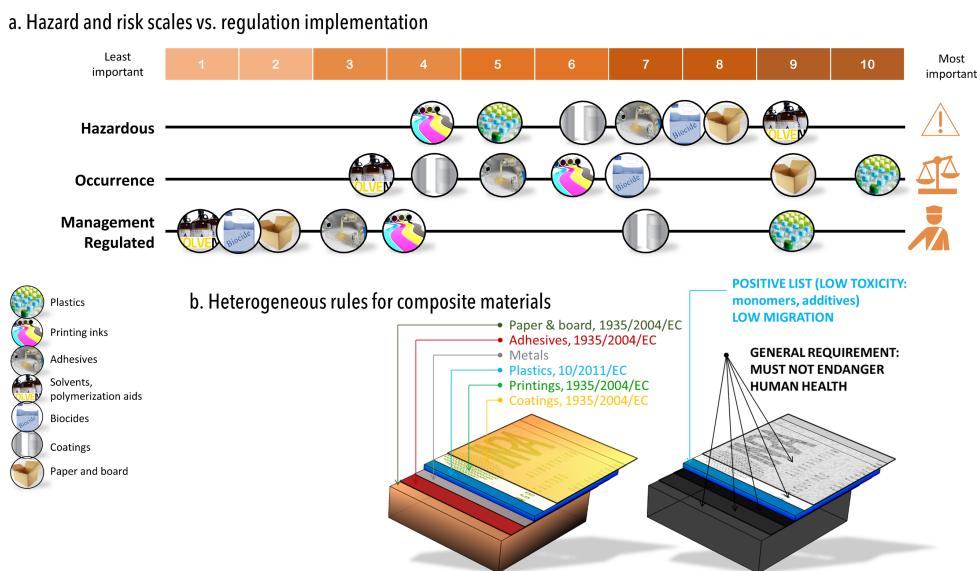


FIGURE 2.1 – Hétérogénéités de principes et de réglementations en Europe pour les différents composants de l'emballage

2.1.3 . Les plastiques : une utilisation et réutilisation très réglementées

La grande majorité des principes (liste positive, critères de pureté, limites de migration spécifiques) ont été repris dans les directives 92/128/CEE et 2002/72/EC puis dans le règlement dit "des matières plastiques" (UE) n° 10/2011 ([EC, 2011](#)) et ses révisions successives. La base légale est l'article 100 du traité de Rome qui recommande une harmonisation de toutes les règles y compris sanitaires qui pourraient avoir une incidence directe sur l'établissement ou le fonctionnement du marché commun. Le symbole du contact alimentaire est obligatoire en 1980 (directive 80/590/CEE). Les racines de la réglementation américaine gérées par la FDA sont similaires et ont pour origine une publication dans Science en 1970 ([Jaeger and Rubin, 1970](#)). La mise en œuvre des règles est toutefois différente et l'administration américaine oppose un droit maximal à exposer le consommateur plutôt qu'un droit à contaminer les aliments. À partir de 2014, la réglementation chinoise se construit sur la base du modèle européen ([2016](#)).

Le concept d'équilibre écologique ou d'équilibre permanent avec le monde naturel n'a pas été repris en tant tel dans les réglementations, mais figure dans le "Green-Deal" européen, la directive (UE) 2019/904 sur les plastiques à usage unique et la loi française n° 2020-105 du 10 février 2020 relative à la lutte contre le gaspillage et à l'économie circulaire, dite "anti-gaspillage" ou AGEC. L'usage de la matière recyclée s'impose même pour le contact alimentaire et notamment parce que les aliments manufacturés sont la première source de déchets de matière plastique : les produits alimentaires ont une durée de vie relativement faibles, ils sont consommés en masse, l'emballage perd sa valeur technique après utilisation. La volonté politique de réduction des déchets plastiques est visible à l'échelle de tous les pays de l'OCDE, mais reste très hétérogène dans les stratégies pour y parvenir ([OECD, 2022](#)). Une solution globale est requise dès que les aliments emballés et leurs déchets circulent entre pays.

a. Food-grade vs. non-food-grade materials



b. Food-grade EU label (official) vs. fancy labels



vs.



c. Recyclability of plastic materials for food contact (closed-loop)

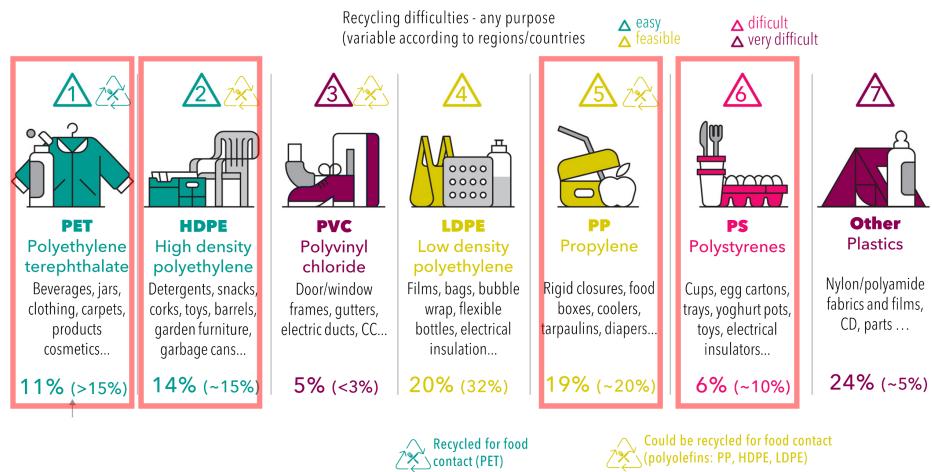


FIGURE 2.2 – Des matériaux plastiques au contact des aliments aux déchets : (a) ils sont mélangés lors de la collecte avec des matériaux plastiques associés à d'autres usages (detergents, cosmétiques...) ; (b) les emballages sont associés à des labels variés certains officiels, d'autres plus fantaisistes car non normés ; (c) la recyclabilité des matières plastiques est très hétérogène en fonction du type de matière plastique.

2.1.4 . Substances non intentionnellement ajoutées

2.1.4.1 . Origine du terme

Le terme de substances non intentionnellement ajoutées (NIAS) a dans un premier temps été utilisé par les industriels pour désigner tout composé chimique issu de contaminations croisées, d'impuretés dans les matières utilisées ou de réactions secondaires non souhaitées. Cette idée d'intention est évoquée dans la littérature pour la première fois par [Gandek et al. \(1989a,b\)](#) à propos d'antioxydants phénoliques pouvant migrer de polyoléfines vers l'eau. Il faudra encore attendre 20 ans avant que le terme de NIAS soit régulièrement utilisé sous cette forme par [Muncke \(2009\)](#) dans un article désignant les emballages comme potentiels réservoirs à perturbateurs endocriniens pour la chaîne alimentaire et devienne récurrent dans la littérature. Ce n'est qu'en 2011 que le terme arrive dans la règlementation européenne (CE) n°10/2011 ([EC, 2011](#)) en incluant les impuretés, les intermédiaires de réaction et les produits de décomposition ou de réaction.

2.1.4.2 . La contradiction environnementale

Choisir entre la protection environnementale, nourrir les hommes et la protection du consommateur : avons-nous le choix ?

Les politiques volontaristes pro-environnementales européennes sont aujourd'hui largement discutées en dehors des frontières européennes et sont à l'origine de controverses sur leurs impacts indirects potentiels à court et moyen termes. Dans le contexte de l'emballage, deux controverses sont particulièrement significatives :

- réduire l'emballage limiterait la disponibilité des aliments, augmenterait le risque de perte des aliments et finalement provoquerait une augmentation des prix particulièrement pénalisante en pleine crise mondiale des matières premières agricoles ;
- généraliser la matière recyclée au contact des aliments provoquerait une augmentation de notre exposition à des substances chimiques dangereuses.

La première controverse est particulièrement active aux États-Unis alors que la seconde l'est en Europe. La seconde controverse sera particulièrement développée dans le reste de l'article, car elle permet de comprendre les garde-fous qui doivent être introduits légalement pour assurer l'obligation d'innocuité des emballages (article 3 du règlement-cadre européen 1935/2004). Les termes de la seconde controverse ont été mis en images sur la Figure 2.3, d'après notamment une circularité de la chimie discutée dans la référence ([Zeng and Li, 2021](#)). Pour l'emballage alimentaire, on retrouvera des éléments du débat dans les discussions européennes autour de la mise en place d'un régime de responsabilité élargie des producteurs ([Hilton et al., 2020](#)) et dans la consultation publique sur la révision de la Waste Framework Directive associées aux décisions 2011/753/EU et (EU) 2019/1004 ([EC, 2022](#)). Le mythe du zéro déchet, le déchet qui devient une matière échangeable et utilisable à l'infini et un consommateur qui consomme des aliments contaminés sont particulièrement illustrés (Figure 2.3a-c). Contrairement à l'intuition, il ne faut pas envisager une situation stationnaire mais un système dynamique où les matériaux et donc les risques sont eux-mêmes en pleine évolution (Figure 2.3d). Le danger proviendrait de l'impossibilité de reconnaître et maîtriser l'origine de la matière et de son mauvais usage (Figure 2.3e).

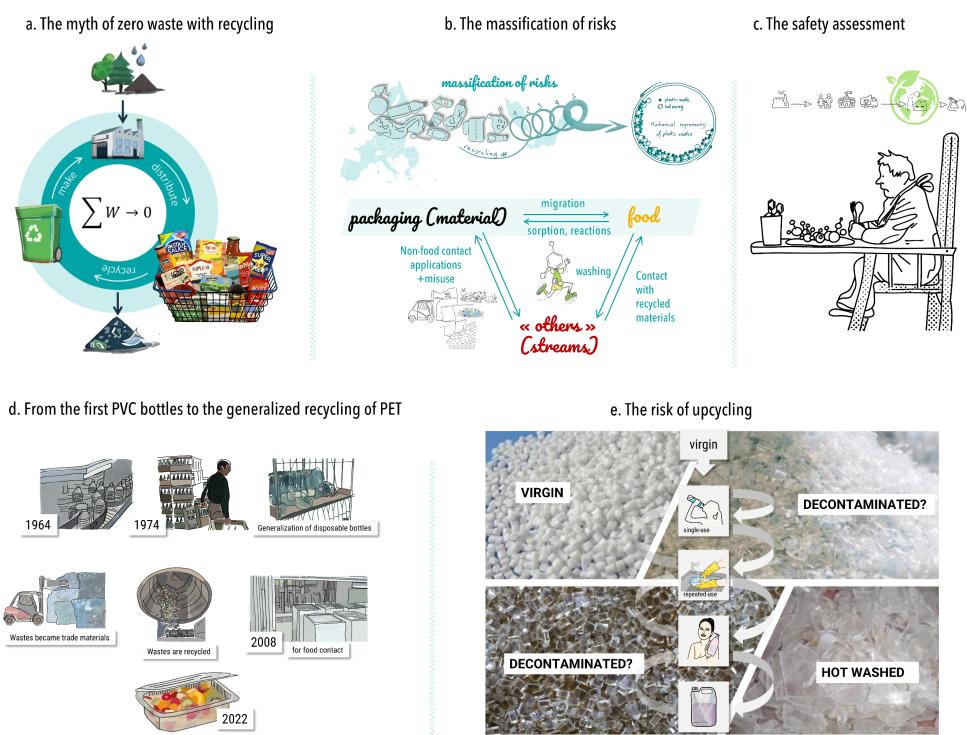


FIGURE 2.3 – *Les termes de la controverse associés à l'intensification de risques de contamination associés à l'utilisation de la matière recyclée pour le contact alimentaire.*

Les matériaux au contact des aliments seraient une des toutes premières sources de contaminants chimiques

Les substances mises en cause dans les crises environnementales (polychlorobiphényles, polychlorodibeno-p-dioxines, polybromobiphényles) et liées à l'emballage (bisphénols, benzo-phénones, diisocyanates de diphenyles et dans une certaine mesure les phtalates) ont de nombreux points communs : homologie structurale, persistance dans l'environnement, effets toxicologiques pouvant être comparables. Dans le cadre d'une économie linéaire, un raisonnement par famille de substances permettait d'adapter les usages aux expositions réelles. Dans le cadre d'une économie circulaire, l'effet est plus insidieux et les contaminants d'une filière pourront se propager à une autre filière. Ils ne pourront être éliminés que par dilution avec des lots moins contaminés ou de la matière vierge. La notion d'exposome et donc d'exposition globale aux substances chimiques est devenue centrale dans les politiques publiques de prévention en France ([ADEME, 2022](#)) et plus généralement en Europe ([Bruinen de Bruin et al., 2022](#)). Les outils pour prédire les perturbateurs endocriniens ([Direction Générale de la Santé, 2022](#)) et suivre les biomarqueurs d'exposition ([WHO, 2019](#)) sont en plein essor. Il s'agit à terme de reconcevoir la chimie à partir de fondations basées sur une meilleure prédition des effets négatifs sur la santé et l'environnement ([Fantke et al., 2021](#)).

Les matériaux présentant un risque pour le consommateur n'auraient pas besoin d'être recyclés

Le polystyrène est le matériau fortement décrié ([Silano et al., 2020](#)) ; c'est le matériau de choix pour les pots thermoformés utilisés pour le conditionnement de produits laitiers (yaourts, fromages blancs, faisselles), de produits pasteurisés conditionnés à chaud (pots de compote). Les données d'exposition des opérateurs par inhalation et les données sur animaux ont conduit le Centre International de la Recherche sur le Cancer (CIRC ou IARC), agence des Nations Unies, à classé son monomère, le styrène, comme agent probablement cancérogène pour l'homme (catégorie 2A) ([IARC, 2019](#)). Le polymère est particulièrement fragile et sensible à une dépolymérisation en présence de contraintes mécaniques et d'oxygène. Cette sensibilité a été proposée récemment comme mécanisme de dépolymérisation à l'échelle industrielle ([Balema et al., 2021](#)), mais est également une source continue de styrène dans les applications du polystyrène. En utilisant les modélisations proposées par [Vitrac and Hayert \(2005\)](#); [Vitrac et al. \(2007\)](#), l'étude de [Vitrac and Leblanc \(2007\)](#) a montré que l'exposition au styrène issu des pots de yaourts était à elle seule comparable à celle des autres sources d'exposition à l'échelle de la population française.

2.1.4.3 . Une formalisation raisonnable

Les risques suscitent émotions et peurs. Ils doivent être expliqués et gérés dans nos démocraties dans le cadre d'approches raisonnées, révisables et consensuelles (Paek and Hove, 2017). La Figure 2.4 pose le problème de la gestion de la matière recyclée. Les méthodes héritées de notre expérience de l'économie linéaire pour assurer la gestion de la sécurité des alimentaires sont disparates et fragmentées suivant la nature des risques (ex. physique, chimique, microbiologique) et l'origine des contaminants (phytosanitaires, matériau au contact, etc.). La nécessité d'une convergence "one food, one health" était déjà soulignée dans les recommandations des Nations Unies publiées en 2003 (FAO, 2003) et sont à la base des standards ISO 22000. La sécurité alimentaire microbiologique est gérée par des approches préventives dont les méthodes HACCP (Hazard Analysis Control Critical Point) et FMECA (Failure Mode Effects and Criticality Analysis) ont été formalisées au début de l'ère spatiale et sont devenues aujourd'hui des standards industriels, à l'échelle de l'atelier à la filière. Pour les matériaux au contact des aliments, la gestion des risques chimiques est bien moins partagée entre industrie chimique et agroalimentaire. Une construction alternative de la sécurité basée sur une sécurité sanitaire construite via les approches FMECA (Nguyen et al., 2013). La rationalité des approches légales appliquées aux matériaux au contact des aliments s'inscrit dans la droite ligne d'un rapport pionnier de l'Académie des Sciences des États-Unis (NRC, 1983), où le danger est associé à une substance et non à un objet ou une pratique. Les limites font l'objet de discussion par les autorités réglementaires et les agences de recherche. Deux points de vue récents sont présentés.

Les limites de la règlementation européenne des matériaux au contact des aliments vues par la Commission Européenne elle-même

Sur la base d'une évaluation de l'implémentation du règlement-cadre européen (CE) n° 1935/2004 par le parlement européen (Karamfilova and Sacher, 2016), l'atelier de la Direction Générale Santé de la Commission Européenne a travaillé en 2021 sur 8 points d'amélioration (EC, 2021a) :

- Manque de fonctionnement du marché intérieur et problèmes de sécurité possibles pour les matériaux non plastiques au contact des aliments.
- Approche basée sur une liste positive de substances non centrée sur la sécurité de l'aliment ou de l'article final.
- Absence de hiérarchisation des substances les plus dangereuses et d'évaluations actualisées.
- L'échange d'informations sur la sécurité et la conformité le long de la chaîne d'approvisionnement est insuffisant et la capacité à garantir la conformité est compromise.
- L'application et la vérification des règles relatives aux matériaux au contact des aliments sont généralement faibles.
- Les règles ne tiennent pas suffisamment compte de la spécificité des petites et moyennes entreprises.
- Les règles n'encouragent pas le développement d'alternatives plus sûres et plus durables.
- Le sujet n'est pas toujours clair et les définitions doivent être revues.

a. Traceability of food vs. a punctual assessment of the packaging



b. New requirements for recycling and obligations hazard identification



FIGURE 2.4 – Comparaison des protocoles de gestion de la sécurité sanitaire pour les produits alimentaires et leurs emballages : (a) sécurité construite par prévention sur les aliments et testée sur produit fini pour les emballages sans traçabilité ;(toutes les bouteilles s'appellent Milk alors que les noms des vaches productrices étaient connus) ; (b) en utilisant des déchets comme matière première des emballages, chaque emballage devient spécifique avec ses contraintes intrinsèques (barrière fonctionnelle, contrôle aux frontières, perturbateurs endocriniens, usage répété) ou imposées au produit alimentaire (durée de vie, possibilité de réchauffage au four).

2.1.5 . Calendrier des évolutions réglementaires

Le calendrier des évolutions réglementaires a été esquissé par les autorités en France, en Europe et en Chine. Il le sera également rapidement aux Etats-Unis. La fin des emballages jetables en plastique est programmée. Tous les contenants en plastique devront être réutilisés en 2040. La situation des autres matériaux mérite clarification, mais ni les mentions biosourcées, biodégradables ou compostables n'ont trouvé gré aux yeux des autorités. Les matériaux biosourcés ne résolvent pas le problème des déchets même s'ils réduisent l'empreinte carbone. Les matériaux biodégradables ou compostables pourraient résoudre le problème des déchets non collectés en permettant de limiter leur impact écologique. Ils conduisent toutefois à une fuite en avant sur l'utilisation de la ressource (énergie et ressource), ils doivent au moins comme les autres être recyclables, réutilisables ou réemployables. L'ensemble doit faire une réduction des emballages primaires, secondaires et ternaires.

2.1.5.1 . Calendrier français vs. calendrier européen

Le réemploi est priorisé, mais il se fera en étapes en passant par une réduction des emballages et le développement du recyclage. Les calendriers français et européens sont comparés sur la Figure 2.5. Ils comprennent à la fois des interdictions de certains usages (vaisselle jetable, barquettes en polystyrène expansé) et des obligations d'utilisation de la matière recyclée. Les objectifs français sont toutefois bien plus ambitieux avec des obligations supplémentaires comme l'obligation d'identifier les perturbateurs endocriniens, l'interdiction des contenants plastiques réchauffables à destination des enfants en restauration collective. L'échéancier laisse peu de temps d'adaptation à l'industrie avec des actions. La bascule sur 100% des plastiques recyclés au 1er janvier 2025 paraît peu réaliste quand globalement un seul plastique est aujourd'hui recyclé pour le contact alimentaire est le polyéthylène téréphthalate (PET). Le marché du PET recyclé est sur les marchés globaux sous très forte tension ([Joltreau, 2022](#)). La rareté de la ressource est responsable du surcoût du PET recyclé qui peut atteindre trois à quatre fois le prix du PET vierge. L'Europe passe rapidement d'un exportateur de déchets plastiques à un importateur net de matière recyclée. Le projet d'amendement des règlements (EU) 1257/2013 et 2020/1056 ([EC, 2021b](#)) vise à encadrer les transports et à tracer les matières contaminées qui ne peuvent pas entrer dans la boucle de recyclage. L'industrie alimentaire s'adapte progressivement à ces nouvelles contraintes en remplaçant le plastique par exemple par de la cellulose moulée. Les applications requérant une stérilisation ou une pasteurisation in situ, un remplissage à chaud, une cuisson sous vide, ne peuvent ni s'accommoder des matériaux cellulosiques, ni du PET, dont la température de ramollissement est voisine de 72°C. Dans les applications de conserve, le plastique (polypropylène) a été proposé comme substitut des vernis époxyphénoliques, notamment pour des boîtes réutilisables.



FIGURE 2.5 – Calendrier des échéances imposées par la loi AGEC française n°2020-105 et la directive européenne 2019/904 relative aux plastiques à usage unique. La partie supérieure montre que les nouvelles restrictions doivent s'accompagner d'une révolution plus profonde.

2.1.5.2 . Les décrets 3R français : obligations et limites

En France, la loi AGEC n°2020-105 propose un contrat d'objectifs de réduction, de réutilisation et de réemploi renouvelés et adaptés par décret pour chaque période consécutive de cinq ans. Pour la période 2021-2025, les objectifs ont été fixés dans le décret n° 2021-517 du 29 avril 2021 ([République Française, 2021](#)). L'article mentionne que "l'objectif de réduction [...] est fixé pour l'ensemble des metteurs sur le marché d'emballages en plastique à usage unique, à 20%, dont au moins 50% obtenus par recours au réemploi et à la réutilisation d'emballages, à l'échéance du 31 décembre 2025, en tenant compte du potentiel propre aux catégories de produits auxquelles sont destinés ces emballages." Il n'y a pas de mention spécifique relative aux applications alimentaires. Des éléments plus techniques sont disponibles dans le document préparatoire du Ministère de la Transition Écologique qui appuie l'évaluation du potentiel de réduction, réemploi et recyclage ([Ministère de la Transition Ecologique, 2020](#)). Les données de CITEO en 2020 ont permis d'établir que les emballages plastiques des produits frais alimentaires représentaient 305 000 tonnes/an et 515 000 tonnes pour les autres produits alimentaires. Mais comme le note le rapport page 18, "la ventilation, au sein de chaque catégorie, selon les différents types d'emballages et de plastiques, n'est pas disponible". Une typologie a bien été proposée par l'ADEME en 2021 pour les couples produit-emballage ([ADEME, 2021](#)), mais elle met peu l'accent sur l'emballage alimentaire qui reste encore peu normé et cantonné à une catégorisation simplifiée des produits frais, secs, boissons, produits de longue conservation... indépendamment des contraintes de transformation, conditionnement, distribution et consommation des aliments, etc. Le concept de matière plastique pour des emballages essentiellement en plastique sans tenir compte des autres usages des polymères thermoplastiques, élastomères, vernis, laques notamment comme enductions. À titre d'exemple, un gobelet en carton destiné à contenir une boisson chaude incorpore un film interne en polyéthylène basse densité (LDPE). Un LDPE à très bas point de fusion sera également utilisé pour coller des parties cartonnées. Afin de guider l'industrie, le Ministère de la Transition Écologique a fait établir des feuilles de route sectorielles (annexe 6 de la stratégie nationale "3R" pour les emballages plastiques à usage unique) ([Ministère de la Transition Ecologique, 2022](#)). Comme indiqué en préambule, ces fiches sont indicatives et ébauchent seulement des cibles à atteindre par secteur en fonction notamment des solutions de remplacement disponibles ou qui le seront rapidement.

2.2 . Les matières recyclées au contact des aliments : une source additionnelle de contaminants

2.2.1 . La pollution de la matière recyclée

La gestion des risques des matières recyclées au contact des aliments est très hétérogène suivant que la matière est en plastique ou d'origine cellulosique. Les papiers et cartons sont recyclés au contact des aliments sans étape de décontamination et d'obligation de traçabilité sur les usages précédents. La boucle de recyclage des papiers et cartons accumule en conséquence les contaminants, par exemple les huiles minérales des encres d'impression. Leur présence a été détectée de manière ubiquitaire au début des années 2000 dans la majorité des aliments secs au contact des papiers et cartons recyclés. Les valeurs très élevées de la contamination sont rapportées dans la Figure 2.6 pour des études ponctuelles (Figure 2.6a) et pour une étude de plus grande ampleur, sur 119 produits commerciaux prélevés dans les supermarchés, réalisée sur la période 2009 et 2011 (Figure 2.6b) par le BfR allemand. Les cinétiques de contamination des aliments et la perte comparable de substance par l'emballage ont été mesurées par le laboratoire cantonal de Zurich. Les résultats exprimés en quantités transférées par rapport à la masse d'huiles minérales dans le matériau démontrent que l'emballage recyclé incluant les parties collées et imprimées sont à l'origine de la contamination observée. Les travaux montrent que la quasi-totalité des huiles minérales (chaînes aliphatiques, photoinitateurs des encres d'impression UV, plastifiants de type phtalates, molécules polyaromatiques) est transférable y compris au travers d'une couche de plastique (Nguyen et al., 2017). Pour les substances les plus petites, leur volatilité est le principal facteur limitant. Pour les substances les plus grosses, la résistance au transfert est dominée par la diffusion dans le film plastique. La quantité maximale transférée dépend seulement de l'affinité chimique pour l'aliment (voir Figure 2.10). Ces observations sont à l'origine d'une évolution de la réglementation française (DGCCRF, 2019) qui introduit plusieurs obligations. La nécessité de prendre en compte le risque de transfert de substances même sans contact, en provenance du suremballage et même au travers un milieu aqueux, est explicitement mentionnée. Les études mentionnées précédemment recherchaient une causalité des transferts pour mieux les prédire. Les études de l'association Food Watch de 2015 (FoodWatch, 2015) et 2021 (FoodWatch, 2021) cherchaient un instantané de la contamination des aliments emballés essentiellement secs sur le marché. Elles démontrent que tous les produits ne sont pas contaminés de la même manière. Les produits à longue durée de vie, secs et/ou gras sont plus contaminés. Les produits pulvérulents comme la semoule sont plus contaminés. La comparaison entre 2021 et 2015 démontre que les pratiques ont évolué : les papiers et cartons recyclés sont aujourd'hui moins utilisés pour le contact alimentaire, la contamination s'en trouve réduite d'autant.

a. Maximum concentrations of mineral oils reported in the literature



Food type	Maximum level of contamination by mineral oils	Origin	Reference
Chocolate and chocolate products	>100 mg·kg ⁻¹	Cardboard	(Lorenzini, 2010)
Baby milk	10-80 mg·kg ⁻¹	Cardboard	(Droz, 1997)
Cereal Products	30 mg·kg ⁻¹	Cardboard and printing inks	(Biedermann, 2013)
Pasta	10 mg·kg ⁻¹	cardboard box	(Biedermann, 2011)
Edible oil	100-1000 mg·kg ⁻¹	Containers	(Wagner, 2001)

b. First food survey of the contamination of mineral oils (2009)

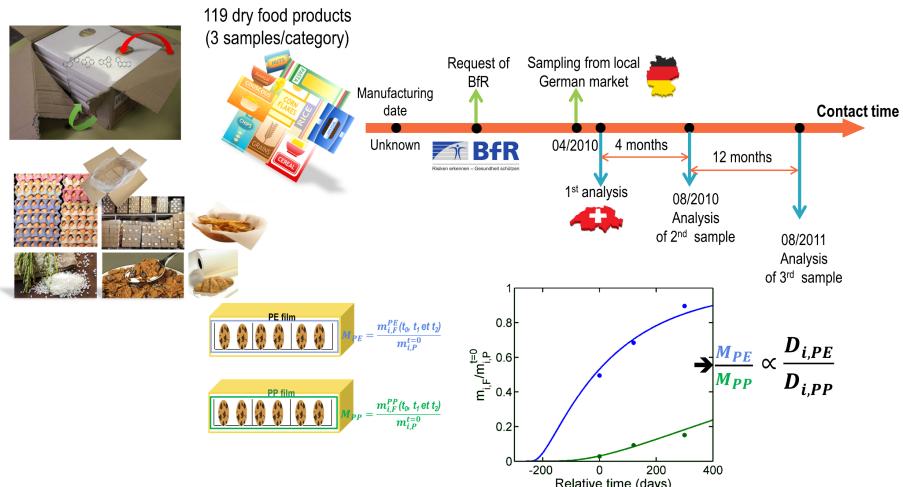


FIGURE 2.6 – Vue générale des niveaux de contamination des aliments par les substances issues des emballages en papier et carton (Lorenzini et al., 2010; Biedermann et al., 2011, 2013).

2.2.2 . Épuration des gisements : une alternative à la décontamination des gisements ?

Les six isomères du bis(isopropyl)naphtalène (DIPN) utilisés comme solvant alternatifs des polychlorobiphényles (PCB) dans les papiers autocopiants sont devenus des contaminants ubiquitaires des papiers et cartons recyclés pour le contact alimentaire. La pollution a été identifiée pour la première fois en 1994 ([Sturaro et al., 1994](#)) avec un risque de contamination des aliments mis en évidence en 2013 ([Lorenzini et al., 2013](#)). Les DIPNs sont des cas particuliers d'huiles minérales aromatiques, qui doivent être considérées comme potentiellement cancérogènes d'après l'Autorité Européenne de Sécurité des Aliments (EFSA) ([EFSA, 2012](#)). La relation entre pollutions croisées lors du recyclage et le risque de contamination des aliments a été particulièrement étudiée pour les matériaux cellulosiques recyclés. On notera une forte implication du milieu associatif. L'Association Food Watch a réalisé deux études en 2015 ([FoodWatch, 2015](#)) et 2021 ([FoodWatch, 2021](#)) à l'échelle Européenne. Les différentes études sont également reprises par le Food Packaging Forum ([Food Packaging Forum, 2017](#)) qui se fait l'écho de tous les risques émergents liés aux nouvelles pratiques ou aux matériaux règlementés de manière spécifique. Il n'existe pas de dispositions générales à l'échelle européenne pour prévenir la pollution en amont des usages pour le contact alimentaire. Seule la France via l'article 112 de la loi AGEC interdit d'utiliser les huiles minérales sur des emballages depuis le 1er janvier 2022. À compter du 1er janvier 2025, il sera interdit d'utiliser des huiles minérales pour des impressions à destination du public. Pour les lettres de prospectus publicitaires et de catalogues non sollicités visant à faire de la promotion commerciale, cette interdiction s'applique à compter du 1er janvier 2023.

2.2.3 . Les limites de la décontamination des gisements

Le basculement d'une économie linéaire à une économie circulaire ne modifie pas fondamentalement la nature des phénomènes, mais introduit une incertitude importante sur la nature des substances transférables. Le socle de substances reste connu et potentiellement maîtrisable dans le cadre d'un recyclage en boucle fermée. Il est très préoccupant si le gisement incorpore de la matière "upcyclée" issue d'un contact non alimentaire (voir Figure 2.3e). La concentration en contaminants dans le gisement peut être réduite par décontamination, mais ils ne peuvent être éliminés. L'explosion du nombre de contaminants potentiels qui en découle est expliquée sur la Figure 2.7. Le scénario proposé tient compte des réactions de dégradation et des contaminations croisées. On retiendra que les processus chimiques tendent à fragmenter les migrants ou contaminants potentiels : les substances sont plus volatiles, plus polaires (cas des produits d'oxydation et d'hydrolyse). Les risques de transferts sont donc accentués. Une analyse des quantités transférées en masse ne traduit ni la diversité des substances, ni le taux de dégradation effectif des polymères et des additifs utilisés.

Le nombre de substances augmente du fait d'une pollution du contenant par son contenu (aliment, produit cosmétique, détergent) lors de la conservation et distribution du produit, de la collecte et du lavage (contamination par les eaux de lavage). Les actes malveillants et les mauvais usages (stockage de produits chimiques) augmentent également les risques pour les consommateurs. Les profils des substances polluantes évolueront également à chaque mise en forme à haute température (ex. extrusion, injection) de la matière recyclée ou quand plusieurs sources de matière seront mélangées. Le cas des multimatériaux est également à considérer de la matière est également échangée dans les assemblages et avec les parties collées et imprimées.

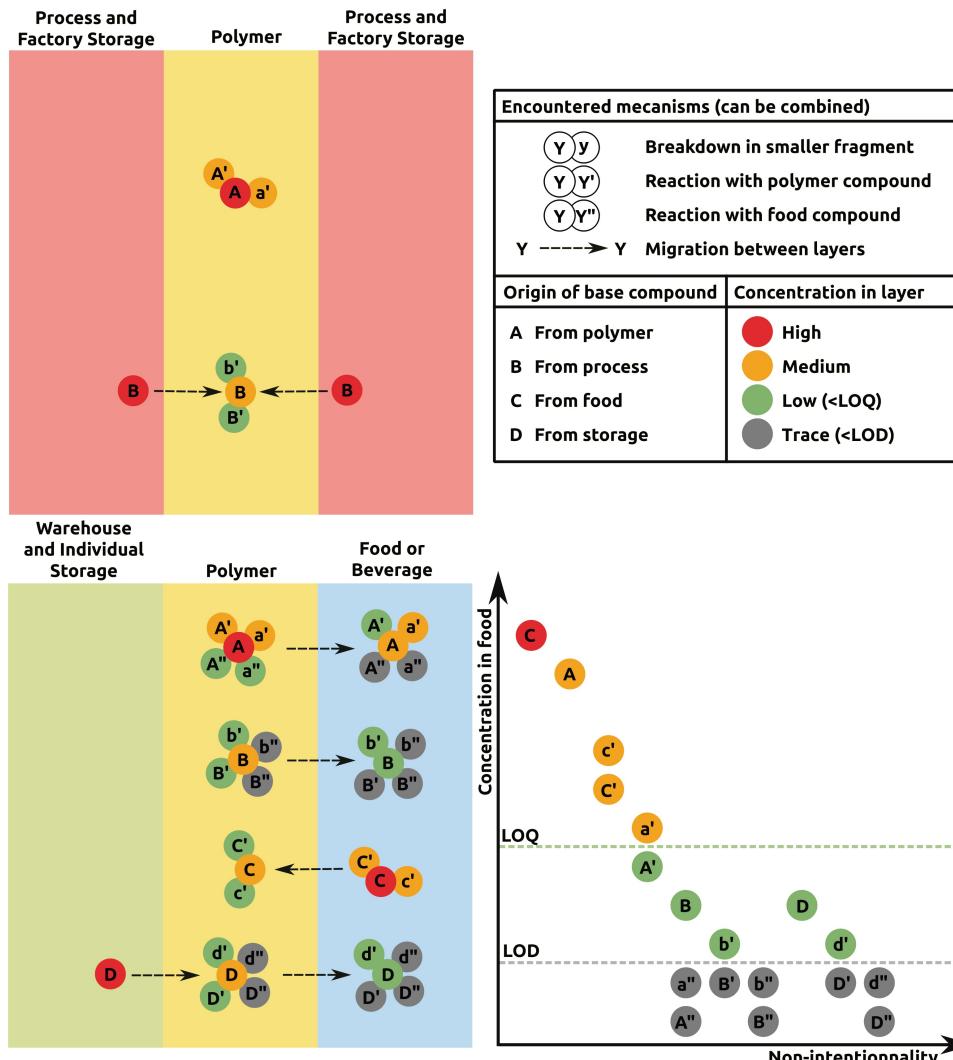


FIGURE 2.7 – La cascade chimique transformant quelques dizaines de substances en des milliers de contaminants potentiels des boucles de recyclage : transferts croisés, décomposition de substances et leurs combinaisons se produisant lors du stockage ou de la collecte des matériaux (en haut à gauche), lors de la mise en contact avec le contenu ou le milieu de lavage (en bas à gauche) et évolution des concentrations attendues dans l'aliment. Les traits en pointillés représentent les limites de quantification (LOQ) et les limites de détection (LOD). L'axe horizontal mesure le caractère non-intentionnel lié à la présence de ces substances. Les apostrophes simples et doubles représentent le degré de transformation des substances par rapport aux substances initiales.

2.2.4 . La gestion de la pollution pour la matière plastique recyclée

Tous les matériaux au contact des aliments doivent ne pas mettre en danger la santé du consommateur et ne pas modifier les caractéristiques organoleptiques des aliments (défauts de goût ou d'odeur). Cette obligation issue de l'article 3 du règlement-cadre européen (CE) n° 1935/2004 s'accompagne également d'une obligation de la mise en place de bonnes pratiques de fabrication (CE) n° 2023/2006, dont les définitions sont laissées à la charge de l'industrie et s'appliquent aux dix-sept groupes de matériaux au contact des aliments (plastiques, élastomères, vernis, liège, colles, encre d'impression, adhésifs, métaux, verre, etc.). Ces exigences imposent donc une obligation de garantir une sécurité chimique équivalente pour la matière recyclée que pour la matière vierge. Pour les matériaux thermoplastiques, le règlement européen (CE) n° 282/2008 (abrogé en septembre 2022 par le Règlement (UE) 2022/1616) a défini les principes des conditions de collecte et de recyclage mécanique (lavage) pour le contact alimentaire. Cette réglementation pionnière a permis le recyclage mécanique des bouteilles PET notamment pour les boissons. Elle s'appuie sur trois piliers :

- des gisements de matière recyclée issus d'un précédent contact pour alimentaire pour au moins 95% d'entre eux ;
- l'homologation des procédés de décontamination qui doivent faire l'objet d'une évaluation par l'EFSA sur leur capacité effective à éliminer des molécules modèles issus de mauvais usages (ex. pollution par du benzène/toluène, lindane, benzophénone...) dont le niveau de contamination a été fixé arbitrairement à 3 mg kg⁻¹ par l'EFSA ;
- la démonstration que la contamination par des substances cancérogènes fortuites ne dépassent pas les seuils de préoccupation toxicologique dans l'aliment (voir le détail dans ([Barthelemy et al., 2014](#))). Ces seuils dépendent du consommateur (bébé, enfant, adulte) et de la toxicité de la substance repris dans le Tableau 2.1.

TABLEAU 2.1 – Valeurs seuils de concentration basées sur l'approche TTC pour un adulte de 60 kg et un apport journalier de 1 kg.

Catégories	Seuil de préoccupation toxicologique (TTC) (en µg/kg p.c. / jour)	C_{thresh}^{TTC} en µg/kg d'aliment (pour 1 kg d'aliment consommé par jour)	C_{thresh}^{TTC} en équivalent LOD*
Avec structures d'alerte pour la génotoxicité	0.0025	0.15	0.015 (ND*)
Organophosphorés et carbamates	0.3	18	1.8 (NQ*)
Cramer classe III	1.5	90	9
Cramer classe II	9.0	540	54
Cramer classe I	30	1800	180

*LOD : limite de détection communément considérée de l'ordre de 10 µg · kg⁻¹, les composés sont non détectables (ND) en deçà de 1 LOD et non quantifiables (NQ) en deçà de 3 LOD

Les trois principes peuvent facilement être mis en défaut dès que l'on s'éloigne des applications eaux de boisson et sodas pour lesquelles les formes de contenants codent leurs utilisations antérieures. Le Tableau 2.2 montre des niveaux supérieurs à $3 \text{ mg} \cdot \text{kg}^{-1}$ pour les contenants en PET issus de détergents et des produits cosmétiques. Pour les boissons alcoolisées ou les applications de type huile et vinaigrette, les niveaux de transferts sont plus élevés et ne permettent pas une utilisation avec 100% de matière recyclée en contact direct avec un aliment. Quand une utilisation de 100% de la matière recyclée n'est pas possible, plusieurs options peuvent être utilisées seules ou en combinaison :

- limiter la durée de vie du produire alimentaire
- ne pas utiliser le contenant à haute température (ex. remplissage à chaud, réchauffage au four à microondes, cuisson sous vide) ; pour le PET, la température de ramollissement se produit à 72°C soit quelques degrés en dessous de sa température de transition vitreuse
- diluer la matière recyclée avec la matière vierge
- utiliser la matière recyclée derrière une couche vierge ou derrière une barrière fonctionnelle.

TABLEAU 2.2 – *Profile des contaminants, classés en fonction de leur masse moléculaire, issus de PET en contact avec des produits non-alimentaires (d'après l'étude de Franz and Welle (2020))*

Substances (Mw g/mol) - N° CAS	Concentration dans un emballage en PET (mg/kg)* en contact avec les produits				
	Liquide vaisselle	Bain de bouche	Produits sanitaires	Shampooing	Savon
Formic acid (46.0)-64-18-6			0.57		
Ethylene glycol (62.1)-107-21-1	3.3	4.17	3.39	3.47	3.67
Tetrahydrofuran (72.1)-109-99-9		0.07	1.06		
n-Butanol (74.1)-71-36-3	0.08		0.04		
Benzene (78.1)-71-43-2	0.15				
Ethanol (78.4)-64-17-5	63.3	40	73.7	185	110
2-Butanone (79.6)-78-93-3	1.25		20.7	0.7	1.86
2-Ethylacrolein (84.1)-922-63-4			0.14		
2-Methyl-1,3-dioxolane (88.1)-497-26-7	0.97	0.87	0.93	1.44	0.89
Toluene (92.1)-108-88-3	0.4				0.49
Hexanal (100.2)-66-25-1			0.41		
2,2-Dimethyl-1,3-dioxolane (102.1)-2916-31-6		1.3			1.11
o-Xylene (106.2)-95-47-6	0.05				
m-Xylene (106.2)-108-38-3	0.1				
Benzaldehyde (106.1)-100-52-7	0.1				1.29
p-Xylene (106.2)-106-42-3	0.02				
Benzylalcohol (108.1)-100-51-6				1.48	1.86
Heptanal (114.2)-111-71-7			0.14		
Benzoic acid (122.1)-65-85-0	0.55				0.96

*Concentration moyenne sur au moins 6 échantillons

Tableau 2.2 – Suite

Substances (Mw g/mol) - N° CAS	Concentration dans un emballage en PET (mg/kg)* en contact avec les produits				
	Liquide vaisselle	Bain de bouche	Produits sanitaires	Shampooing	Savon
Phenoxymethanol (124.1) or Salicylalcohol (2-hydroxymethyl phenol) (124.1)-90-01-7	2.2				
Octanal (128.2)-124-13-0			0.27		
1-Butoxy-2-propanol (132.2)-5131-66-8			4.57		
1-Butoxypropan-2-ol (132.2)-5131-66-8			0.33		
Limonene (136.2)-7705-14-8	0.38	0.07	2.67	0.48	1.64
4-Methoxy benzaldehyde (136.2)-123-11-5			0.27		
2-Hydroxy-4-methyl benzaldehyde (136.2)-698-27-1	0.07				
1-Phenoxy ethanol (138.2)-122-99-6	1.4				
2-Menthene (138.3)-5113-93-9. 2230-85-5		0.07			
1-Menthene (138.3)-5502-88-5		0.03			
Acetic acid hexyl ester (144.2)-142-92-7 (n) or 628-95-5 (iso)	0.12				
Nonanal (142.2)-124-19-6	0.2		0.41		
cis-Anethole (148.2)-25679-28-1		1.27	4.43		
trans-Anethole (148.2)-4180-23-8		11.5	6.28		
Carvone (150.2)-99-49-0		0.4			

*Concentration moyenne sur au moins 6 échantillons

Tableau 2.2 – Suite

Substances (Mw g/mol) - N° CAS	Concentration dans un emballage en PET (mg/kg)* en contact avec les produits				
	Liquide vaisselle	Bain de bouche	Produits sanitaires	Shampooing	Savon
Acetic acid benzyles- ter (150.2)-140-11-4	0.43		0.46	0.52	
Methyl salicylate (152.1)-119-36-8		0.23			
Methyl parabene (152.2)-99-76-3		0.1			
2-(Benzyl)ethanol (152.2)-622-08-2					0.29
Menthone (154.3)-89-80-5		1.17	0.26		
iso-Menthone (154.3)- 1196-31-2. 491-07-6		0.22	0.09		
Decanal (156.2)-112-31-2	0.17				
Menthol (156.3)-89-78-1		1	0.41		
iso-Eugenol (164.2)-97-54-1 or Eugenol (164.2)-97-53-0		0.13			
Anisketone (164.2)-122-84-9		1.23	8.4		
2-Hydroxybiphenyl (170.2)-90-43-7				0.3	
Hexadecene (224.4)-629-73-2					0.2
Inconnus			[0.13-5.75]		

*Concentration moyenne sur au moins 6 échantillons

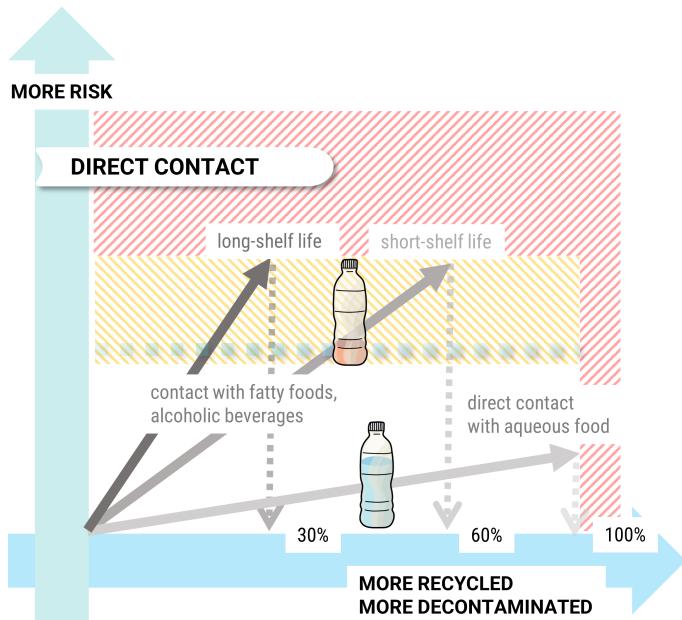
2.2.5 . La barrière fonctionnelle : un concept pour maximiser la matière peu ou pas décontaminée

Le concept de barrière fonctionnelle a été introduit dans le règlement (EU) n° 10/2011 pour permettre l'utilisation de substances non évaluées dans les matières plastiques, dont des contaminants sauf s'ils sont cancérogènes, mutagènes et reprotoxiques (CMR) ou sont sous forme nanométrique. La barrière fonctionnelle est un concept un peu flou, où un matériau est jugé suffisamment barrière pour ne mettre en danger la santé du consommateur. Le caractère flou de la barrière provient du fait que la barrière n'est pas absolue, mais éphémère, et parce que les propriétés barrières dépendent de la substance et de la température. Dans le contexte actuel, elle permettrait d'utiliser de la matière peu ou pas décontaminée. Elle offre potentiellement une alternative à la dilution de la matière recyclée par de la matière vierge. La barrière peut être constituée d'un même matériau vierge ou recyclé et fortement décontaminé ou d'un autre matériau, incluant un traitement de surface de type oxyde de silice ou de carbone amorphe. La Figure 2.8 illustre le principe dans une logique d'optimisation sous contrainte : maximisation de la matière recyclée tout en garantissant une sécurité sanitaire et une durée de vie de produit équivalentes. Le concept de barrière s'impose dès que le risque de contamination est important pour l'aliment du fait d'une forte affinité chimique des contaminants pour l'aliment (aliments gras, boissons alcoolisées), temps de contact longs et/ou à haute température, ou encore de l'utilisation de matériaux fortement contaminés.

a. Comparison of recycled materials used without and behind a functional barrier



b. Recycled material mixed with virgin one in direct contact with food



c. Recycled material used behind a functional barrier

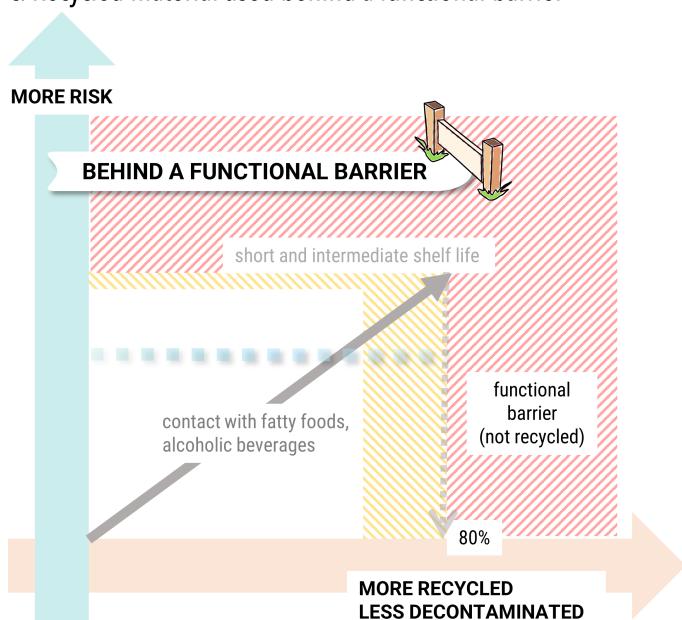


FIGURE 2.8 – Illustration de l'intérêt de l'utilisation d'une barrière fonctionnelle dans le design d'un emballage incorporant de la matière recyclée. Le 100% matière recyclée en plastique est essentiellement applicable aux eaux minérales et sodas. Pour les autres applications, une dilution de la matière recyclée ou une barrière fonctionnelle sont requises.

2.2.6 . La difficulté du contrôle et de l'orientation des flux de matière recyclée

Les scénarios présentés tout comme la mise en situation présentée sur la Figure 2.9 mettent en évidence la difficulté des contrôles par les autorités et la possibilité d'assurer la traçabilité des matériaux une fois que ceux-ci seront introduits dans la boucle de recyclage. Comment maîtriser la teneur en matière recyclée dans les gisements notamment si celle-ci est sandwichée entre deux couches de vierge ou mélangées ? Comment prévenir l'accumulation de contaminants non alimentaires dans la boucle de recyclage, notamment si de la matière peu décontaminée (donc moins chère) ou même non homologuée serait utilisée en couche éloignée de l'aliment. Ces questions ont été étudiées récemment dans le cadre d'un programme de recherche collaboratif pour les barquettes thermoformées de PET de type ABA, où A est une couche recyclée décontaminée et B une couche peu ou pas décontaminée. La contamination des couches A par la couche B se produit dès la mise en forme de la plaque de PET (coextrusion) ou de sa mise en forme (thermoformage) et se poursuit lors du stockage du matériau même en l'absence d'aliment. Ces phénomènes ont été étudiés dans un cas plus général avec un contenant et son suremballage par Nguyen et al. (2019). Le problème qui existe pour le PET (30% des applications alimentaires) se répète de manière plus aigüe pour les autres matrices polymères aujourd'hui peu ou pas recyclées en boucle fermée et présentant un potentiel important : polyéthylène haute-densité et polypropylène. Parce que ces matrices sont caoutchoutiques à température ambiante, elles sont plus facilement chargées en polluants et en polluants de plus forte masse moléculaire. Les niveaux de concentration en polluants à considérer sont typiquement de 1000 mg kg⁻¹.

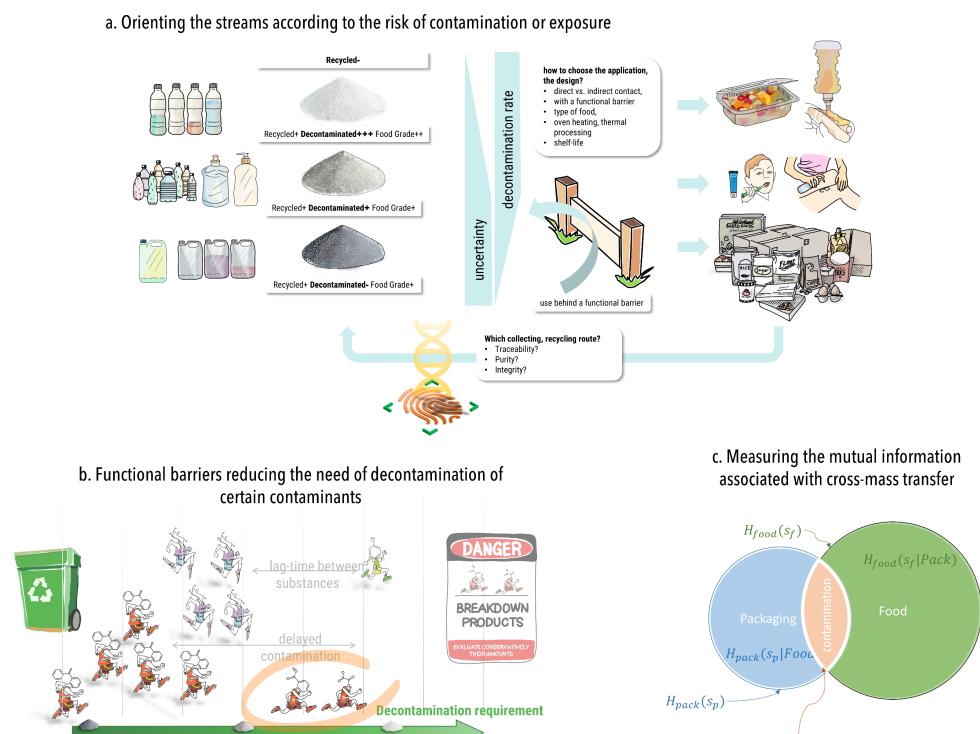


FIGURE 2.9 – Complexité du contrôle et de l'orientation des flux illustrés par des exemples. (a) Orientation des flux en fonction de l'origine et du niveau de contamination des gisements. (b) Lissage du besoin de décontamination de la matière recyclée en fonction de la vitesse de diffusion des substances (la matière recyclée pourrait être plus riche en contaminants de haut poids moléculaires). (c) Représentation de la contamination croisée entre le contenant et son contenu sous la forme d'entropie mutuelle échangée ici par les compartiments emballage et aliment.

Depuis le 10 octobre 2022 et pour tenir compte de la nécessité de nouveaux procédés de recyclage (lavage par solvants, dépolymérisation partielle ou totale), le règlement (CE) n° 282/2008 a été abrogé et remplacé par le règlement (EU) n° 2022/1616 à la fois plus général et permettant une utilisation de nouvelles technologies avant homologation finale. De nouvelles dispositions sont introduites pour améliorer la maîtrise de la traçabilité et de la qualité de la matière recyclée :

- la définition d'un lot "une quantité de matière de même qualité, produite moyennant l'application de paramètres de production uniformes lors d'une certaine étape de fabrication, stockée et conservée de manière à exclure tout mélange avec d'autres matières ou toute contamination"
- "traçabilité de chaque lot jusqu'au premier tri des déchets plastiques collectés" et "certification par un tiers indépendant."
- "Un recycleur [...] surveille le niveau moyen de contamination en s'appuyant sur une stratégie d'échantillonnage fiable qui échantillonne les lots de matières premières plastiques et les lots décontaminés produits correspondants. La stratégie d'échantillonnage tient compte de tous les facteurs susceptibles d'avoir une incidence sur la composition des matières premières plastiques et s'intéresse en particulier aux variations dans l'origine de ces matières, qu'elles soient géographiques ou d'une autre nature."
- En particulier, il est demandé "une liste de toutes les substances présentant une masse moléculaire inférieure à 1 000 daltons détectées dans les matières premières plastiques de chacune des installations de décontamination ainsi que dans les matières plastiques recyclées issues de ces installations, triées par ordre décroissant de leur présence relative et dont au moins les 20 premiers contaminants fortuits détectés dans la matière première ont été identifiés, les quantités étant exprimées sous la forme d'une fraction massique de la matière première et de la matière produite,"
- "une analyse de l'origine la plus probable des contaminants identifiés visés aux points c) et d) ainsi qu'une analyse visant à déterminer si ces origines pourraient donner lieu à la présence simultanée d'autres substances préoccupantes non détectées ou non identifiées au moyen des techniques d'analyse appliquées ;"
- "une mesure ou une estimation des niveaux de migration vers les denrées alimentaires des contaminants présents dans les matériaux et objets en matière plastique recyclée ;".

2.2.7 . Ce que nous apprennent les crises passées sur les risques futurs : décalque, transferts sans contact

Des phénomènes peu ou pas compris sont à l'origine de crises régulières au hasard des progrès de la détection des contaminants dans les aliments. Ils sont dans la plupart des cas associés à un ou des raccourcis empruntés par les molécules et que l'industrie n'a pas su anticiper. Le prototype de ces raccourcis est le phénomène de décalque qui a été mis en évidence en 2005 lors de la crise de l'isopropyle thioxanthone ou ITX, un photoinitiateur utilisé pour les encres UV ([Gillet et al., 2009](#)). L'enroulement du film des briques Tétrapack mettait en contact les parties imprimées avec la parties internes de l'emballage. Le transfert n'était pas possible du fait de la présence d'une couche d'aluminium interne au film mais possible par décalque. Des phénomènes très similaires se produisent quand la matière recyclée est placée derrière une barrière dite fonctionnelle et que le matériau est stocké en rouleaux, plaques empilées ou quand les articles finis (pots, barquettes) sont emboîtés. Les possibilités de transferts au moment de la mise en déchets, collecte, du lavage et recyclage mécanique sont très nombreuses et sont insuffisamment caractérisées. Les modes de transfert déjà identifiés sont listés dans le Tableau [2.3](#). Les substances peuvent être transférées lors d'un contact temporaire entre deux matériaux présence de forces extérieures (friction, pression), d'un lavage ou d'un transport par advection. Sinon, les transferts entre matériaux restent lents, mais restent probables dès que les temps de contact sont ou de partage de voisinages sont prolongés ou soumis à des températures élevées. Les transferts peuvent ainsi se produire en cascade au travers plusieurs matériaux comme cela a été montré pour la contamination des milieux de culture par les monomères des suremballages des réacteurs à usage unique utilisés par l'industrie biopharmaceutique ([Nguyen et al., 2019](#)). Les transferts peuvent également se produire au travers d'autres matériaux comme cela est montré sur la Figure [2.10](#) pour le transfert des substances du suremballage en carton vers l'aliment au-travers de deux films d'air et d'un film de polypropylène biorienté de 50 µm d'épaisseur. Tout matériau de réception peut devenir à son tour matériau source. Il n'a pas besoin de simultanéité pour que la cascade de transferts se poursuive. Au cours du temps, les quantités transférables sont censées être réduites par les multiples diffusions. Cet effet est toutefois négligeable si le volume du réservoir source est grand devant la ou les phases de réception. Dans le cas des procédés de recyclage mécanique, les colles, les encres et les substances biocides accumulées dans les eaux de lavage représentent un réservoir important de substances. Le lavage du polyéthylène téréphtalate à une température supérieure à sa température de transition vitreuse (76-81°C pour le PET sec et jusqu'à 60°C pour le PET plastifié par l'eau) est favorable à une absorption rapide des contaminants par un plastique qui devrait être lavé et non recontaminé. Le même mécanisme est l'œuvre quand les contenants sont lavés en présence de résidus de contenus, dans ce cas les contaminants sont les substances des produits traités : détergents, produits cosmétiques, matériaux de construction...

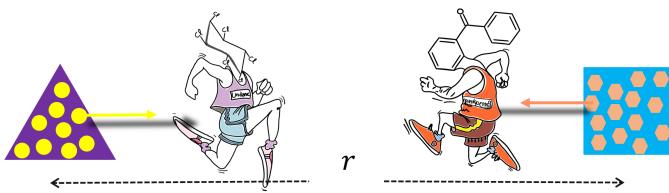
TABLEAU 2.3 – Mécanismes de transferts croisés entre matériaux capables de redistribuer les contaminants dans la filière de recyclage.

Mécanisme	Source	Réception	Mode de transfert	Facteurs activant le transfert
Décalque ou set-off	Surface ou matériau enduit ou imprimé	Tout matériau plus “sec” sur la source en contact temporaire ou prolongé avec la source	Abrasion mécanique (transfert rapide)	Friction, pression
			Entraînement par capillarité (transfert relativement rapide)	Présence d'un liquide (ex. solvant d'impression), fusion partielle du solvant solide de type cires (lors de l'entreposage, transport à haute température), lavage à l'eau chaude...
			Diffusion de surface ou en phase vapeur (transfert lent)	Temps de contact, température
Lavage de la surface	Surface imprimée, collée	Tout matériau poreux	Entrainement par l'eau ou solvant et séchage	Température, présence de détergents (soude)
Transfert sans contact au travers d'une couche d'air non renouvelée	Tout matériau contenant des substances	Tout matériau placé dans un voisinage proche de la source en présence d'une convection limitée	Transfert par déplacement de l'équilibre entre le solide et l'air : désorption et diffusion dans la phase gazeuse suivie d'une réabsorption dans la phase de réception	Temps de contact longs avec une activation très importante par la température. Le transfert est d'autant plus favorable que le matériau de réception présente une grande surface d'échange et a une affinité importante pour la ou les substances

Tableau 2.3 – Suite

Mécanisme	Source	Réception	Mode de transfert	Facteurs activant le transfert
Transfert sans contact au travers d'un film liquide	Tout matériau contenant des substances	Tout matériau placé dans un voisinage proche ou mélangé avec la source dans un récipient/conteneur/réacteur agité	Transfert par déplacement de l'équilibre entre le solide et le liquide : désorption et diffusion ou advection dans la phase liquide suivie d'une réabsorption dans la phase de réception	L'agitation et l'absence de renouvellement du liquide sont favorables au transfert.
Mélange de matière	Fraction contaminée	Fraction non contaminée	Tous les mécanismes précédents et diffusion moléculaire dans le cas d'un mélange à l'état fondu (cas des thermoplastiques) ou de l'utilisation d'un solvant	La contamination de la matière est immédiate et proportionnelle au taux d'incorporation de la matière "contaminée". Ce mécanisme prévaut pour les métaux lourds.

a. Cross-mass transfer between materials without contact



b. Mass transfer of mineral oils from the outer packaging to a model dry food across its plastic packaging

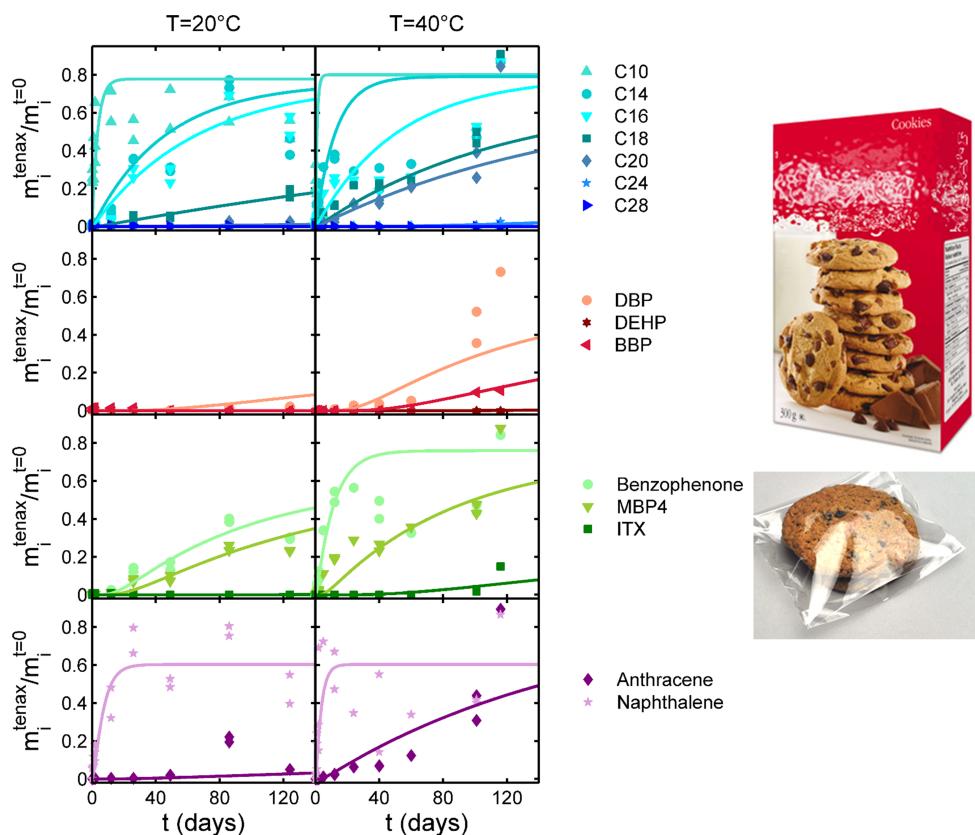


FIGURE 2.10 – Exemple de transfert de matière entre matériaux et de contamination de l’aliment sans contact (Nguyen et al., 2017) : (a) principe général des transferts croisés qui conduit à uniformisation des contaminants à grande échelle ; (b) cinétique des quantités relatives transférées dans l’aliment en provenance du suremballage pour un produit sec de type cookie au travers d’un film polypropylène biorienté de 50 µm d’épaisseur.

2.3 . Apport des mathématiques à l'analyse des sources et mécanismes de contamination

Les contaminations ou plus exactement les migrations (transferts) sont constatées sur les emballages et les articles fini. Dans la plupart des cas, elles sont jugées acceptables si elles sont inférieures à un seuil règlementaire ou toxicologique. Le contrôle tel qu'il est réalisé dans la réglementation ne prend pas en compte les possibilités d'évitement et de réduction des risques de transfert. Dans la cadre d'une économie circulaire, les risques devront être gérés de manière préventive de manière à minimiser l'exposition à des substances non évaluées ou inconnues. Le concept de prévention a été discuté par Nguyen et al. (2013) dans la logique de la conception d'emballages sûrs. L'outil principal de l'approche proposée est la modélisation mécaniste des transferts. Elle n'est possible que si l'identité des substances est connue (Gillet et al., 2009) ou si son occurrence est définie dans un cadre probabiliste (Vitrac and Hayert, 2005). Cette partie discute des possibles contributions des mathématiques à un problème de contamination des aliments dans le cadre d'une économie circulaire où domine l'incertitude tant sur la nature, l'occurrence et la quantité des substances.

2.3.1 . La modélisation de la migration pour l'analyse des causes et des effets

La modélisation de la migration a été proposée à la fois aux États-Unis (Limm and Hollifield, 1996) et dans l'UE (Baner et al., 1994, 1996; Hinrichs and Piringer, 2002) pour l'évaluation de la migration. Elle est aujourd'hui étendue dans l'UE via le groupe de travail spécifique TF-MATHMOD publant des guides pratiques (Hoekstra et al., 2015). La modélisation de la migration est une avancée majeure qui a permis de gérer la sécurité des matières plastiques dans le cadre d'une économie linéaire (test de conformité) et également dans le cadre d'une économie circulaire indépendamment de la capacité à mesurer. Dans ce contexte, elle est un élément de compétitivité qui a été accueilli très favorablement en Europe, aux Etats-Unis et en Chine qui l'ont adopté comme principe général. Les principes généraux sont repris dans la Figure 2.11 et dans le Tableau 2.4 ainsi que discutés en détails dans la référence (Zhu et al., 2019b). Elle peut être également vue comme une étape de la conception assistée par ordinateurs (ou computer-aided food manufacturing) récemment revue par un collectif d'universitaires européens et américains (Datta et al., 2022). La virtualisation des tests et la mathématisation des décisions peuvent être poussées pour intégrer les obligations environnementales : minimiser les déchets et la ressource, maximiser la durée de vie de l'aliment pour réduire la perte de déchets, maximiser la matière recyclée, etc. C'est sur base que le logiciel FMECAengine (Nguyen et al., 2013) a été généralisé (Zhu et al., 2019a) pour permettre la conception de bouteilles 3D (2/3 des déchets) pour des applications critiques.

Il n'y a pas de limite à la portée de la modélisation dans les réglementations et pratiques actuelles. Les tests de conformité peuvent être considérés comme l'utilisation la plus simple de la modélisation et la validation du recyclage en boucle fermée comme l'application la plus complexe. L'utilisation répétée, les matériaux composites et les approches de sécurité dès la conception peuvent être envisagées avec une complexité intermédiaire. Les modèles antérieurs ne pouvaient couvrir que des matériaux uniques et des géométries simples sans aucun changement dynamique des conditions. Les modèles les plus avancés peuvent aujourd'hui intégrer des informations à l'échelle moléculaire et couvrir toute une chaîne d'approvisionnement.

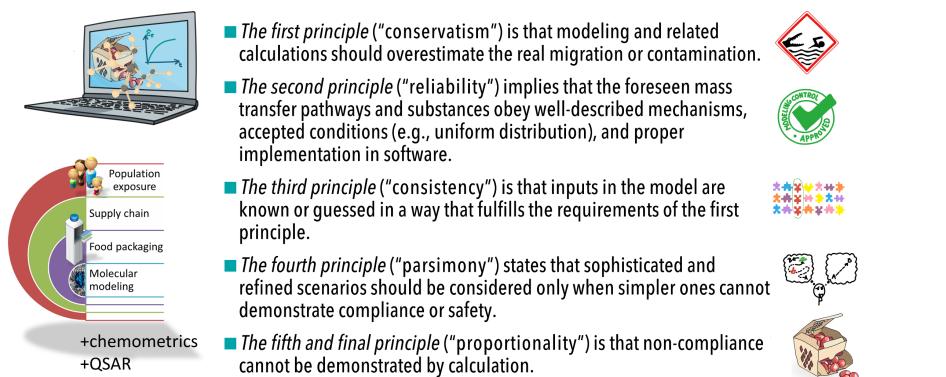
La valeur exacte de la contamination des aliments n'est jamais atteignable, car les conditions de contact sont variables (temps, température) et la connaissance des mécanismes moléculaires

TABLEAU 2.4 – Résumé des principaux concepts de la modélisation de la migration

Type	Description
Worst-case	Approximation de 1er niveau, la plus grande marge de surestimation : transfert total de substances issues des matériaux aux aliments. La concentration correspondante dans l'aliment est déterminée par le bilan de matière.
Hypothèses conservatrices (Hoekstra et al., 2015)	Application pour la démonstration de conformité avec une surestimation de la concentration dans l'aliment, le calcul ne peut pas démontrer une non-conformité.
Évaluation de l'étape critique (Nguyen et al., 2013)	<ul style="list-style-type: none"> - Adaptation de l'approche FMECA (Failure Mode Effects and Criticality Analysis) aux transferts de matière pour identifier les composants de l'emballage, les substances initiales des polymères ainsi que les étapes de la chaîne de valeur qui sont critiques : méthodologie de conception des emballages sûrs (Nguyen et al., 2013). - Construction d'une échelle de sévérité exprimant le rapport entre la concentration dans l'aliment (C_F) et un seuil acceptable (T).
Probabiliste et d'évaluation d'exposition (Vitrac and Hayert, 2005, 2007 ; Vitrac et al., 2007 ; Vitrac and Leblanc, 2007)	<ul style="list-style-type: none"> - Au contraire de la modélisation classique qui calcule une valeur de concentration dans l'aliment, la modélisation probabiliste calcule la distribution statistique des quantités de substances transférées pour toute combinaison de paramètres (probable ou non). - Les effets combinés de l'incertitude et de la variabilité (par exemple, les temps de séjour à des températures spécifiques) peuvent être analysés ensemble et utilisés pour fournir des estimations prudentes avec un risque contrôlé de sous-estimation. - La même approche peut identifier les paramètres les plus influents agissant sur la valeur de la migration ou la valeur de décision finale.

n'est pas parfaite. En conséquence, la migration est estimée via une modélisation itérative (tier-modeling) avec des approximations successives de la migration (voir Figure 2.11b). Au premier niveau, l'estimation est très grossière et liée au facteur de surestimation le plus élevé. Si la concentration déterminée au niveau n est supérieure au seuil préoccupant, le niveau suivant est déclenché en introduisant des améliorations et des détails substantiels, et ainsi de suite. Le processus s'arrête lorsqu'aucune information supplémentaire ne peut être introduite (les expériences doivent être préférées) ou lorsque la concentration calculée est inférieure au seuil préoccupant. Le niveau le plus bas du seuil de préoccupation définit le niveau approprié de connaissances requis pour démontrer la conformité ou pour garantir l'utilisation sûre d'un matériau, d'une substance ou d'un processus. Il n'y a pas de procédure systématique pour identifier le niveau minimum pour atteindre l'objectif, et seules les informations nécessaires peuvent être listées. Dans un futur proche et dans le cas de l'évaluation de la matière recyclée, l'évaluation doit rester réaliste et prendre en compte l'usage attendu de la matière recyclée. Il s'agirait d'intégrer plusieurs variabilités, notamment les pratiques des ménages : niveaux de consommation, fréquences des achats qui contrôlent avec le volume des achats le temps de contact entre l'emballage et son aliment.

a. The five principles of migration modeling



b. Tier modelling to demonstrate compliance and safety

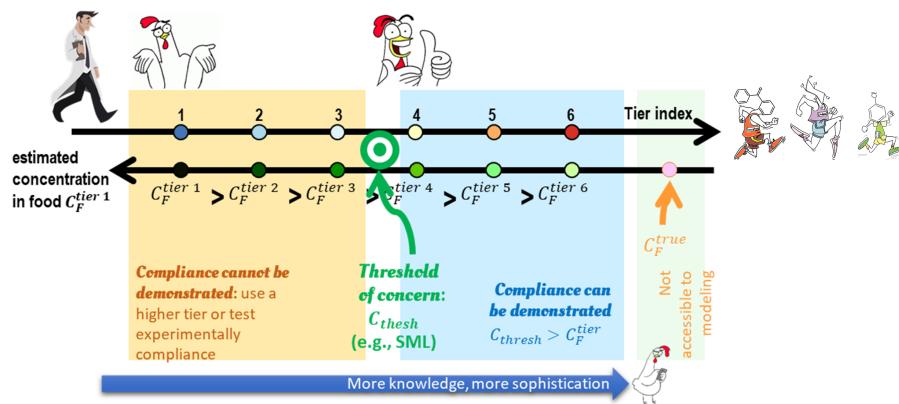


FIGURE 2.11 – Principes reconnus (voir ([Hoekstra et al., 2015](#))) et méthodologie pour démontrer la conformité des matériaux au contact des aliments via une modélisation itérative (tier-modeling) utilisant un enrichissement et raffinement des hypothèses en fonction des résultats obtenus à l'étape précédente (pour un tier index plus faible). La conformité est démontrée dès que la concentration estimée est supérieure au seuil de préoccupation (). Le tier index 1 est généralement associé à une hypothèse de migration totale (la totalité de la substance est transférée à l'aliment) et on consomme quotidiennement un kilogramme de cet aliment contaminé au même niveau. Le règlement (EU) 10/2011 n'impose pas une méthode de calcul ou une limite dans le détail des hypothèses reconnues. L'industrie chimique qui a accès aux détails de ces matériaux pourra utiliser des approches très détaillées comme la modélisation moléculaire, alors que l'industriel agro-alimentaire et le transformateur de matière utiliseront des scénarios extrêmement conservatifs.

2.3.2 . La recherche de la signature chimique des contaminants

La recherche de la signature chimiques de substances dans un extrait d'un emballage ou dans un migrant donne lieu à de nombreuses variations du problème mathématique de déconvolution. Le terme déconvolution est utilisé de manière abusive et doit être compris comme une factorisation d'une fonction (ou signal mesuré) en une somme de sous-fonctions (ou signaux reconstruits). Parce qu'il n'y a pas d'opération de convolution, le terme de "désommation" serait plus adapté. Les inconnues sont le nombre de sous-fonctions, leurs poids et éventuellement les sous-fonctions elles-mêmes. L'occurrence chimique d'une substance est mathématiquement démontrée si la signature correspond soit à son empreinte (ressemblance) ou si son poids est corrélé à sa concentration. La déconvolution dite aveugle utilise les propriétés statistiques des moments d'ordre supérieurs à 2 pour reconstruire des sous-signaux qui vérifient des propriétés notamment d'indépendance (Hyvärinen, 2013). Les contraintes associées aux différentes techniques sont nombreuses : constitution préalable d'un dictionnaire de sous-signaux, nombre d'échantillons disponibles, nature physique ou non des sous-signaux, possibilité de décorrélation, taille de l'espace chimique à étudier. Dans la logique d'une économie circulaire, le nombre de contraintes ne peuvent pas être vérifiées. Cette partie discute les principes et les limites des techniques actuelles.

2.3.2.1 . Le concept de source physique

Le concept de source physique est une manière de décrire un échantillon comme étant un mélange physique de corps moins complexes appelés sources, ces sources peuvent être un simple composé isolé tout comme un mélange en soi. La nature simple ou complexe des sources à considérer dépend d'une part du besoin subjectif de précision pour décrire un échantillon, mais aussi du cas d'étude pouvant comporter des composés exogènes à tout mélange possible. Les types de cas d'étude identifiés et pouvant influencer la caractérisation des sources sont résumés dans la table 2.5.

TABLEAU 2.5 – Cas d'étude types pour le concept de sources

Cas d'étude	Décomposition en sources
Mélange de deux solutions S_1 et S_2 connues pour obtenir S_m	Le mélange peut s'écrire facilement avec un facteur de proportion α : $S_m = \alpha S_1 + (1 - \alpha) S_2$
Ajouts dosés de N molécules ou mélanges de molécules M_i dans une solution S pour obtenir S_a	La description du mélange nécessite dans le cas général $N + 1$ sources : $S_a = S + \sum_{i=1}^N \beta_i M_i$
Mélange inconnu de deux gisements G_1 et G_2 comportant respectivement N_1 et N_2 molécules avec recouvrement possible $N_{1\cap 2}$ pour obtenir G_m	Les gisements G_1 et G_2 n'étant pas accessibles, la description doit se faire avec les $N_1 + N_2 - N_{1\cap 2}$ molécules M_i comme sources : $G_m = \sum_{i=1}^{N_1+N_2-N_{1\cap 2}} \gamma_i M_i$

Nous pouvons voir finalement que le problème de convolution de sources n'est pas aussi trivial et univoque qu'en apparence, le problème inverse s'avère encore plus compliqué à cause du manque d'informations sur le nombre de sources composant un mélange ou un signal et sur la nature de ces sources. Afin d'appuyer notre réflexion, nous allons développer ce concept de sources dans le cas de chromatogrammes.

Nous allons considérer un chromatogramme $\mathbf{Y}_{(M,T)}$ représenté par une matrice de taille M par T , chaque ligne représentant une valeur de ratio masse sur charge et chaque colonne un temps de rétention dans l'analyse. Dans le cas de nos analyses les matrices sont constituées de 611 lignes et environ 13000 colonnes.

$$\mathbf{Y}_{(M,T)} = \begin{pmatrix} y_{1,1} & \dots & y_{1,T} \\ \vdots & \ddots & \vdots \\ y_{M,1} & \dots & y_{M,T} \end{pmatrix}$$

Afin de simplifier les notations, nous allons introduire la notation \mathbf{Y} pour désigner le chromatogramme déplié par colonnes $\mathbf{Y} := \mathbf{Y}_{(M,T)}$.

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} \\ \vdots \\ y_{M,1} \\ y_{1,2} \\ \vdots \\ y_{M,2} \\ \vdots \\ y_{1,T} \\ \vdots \\ y_{M,T} \end{pmatrix}$$

La convolution est une opération en apparence simple d'addition de signaux avec des poids variables, mais qui devient complexe à partir du moment où le concept est appliqué à un système physique analysé par une quelconque technique analytique. En effet la convolution a lieu au moment du mélange physique en amont de l'analyse, la linéarité du problème et l'identité des sources n'ont en général aucune garantie d'être conservée dans le signal. La déconvolution décrit l'opération inverse et permet de décomposer un signal complexe en somme de signaux plus simples, encore une fois l'application d'une telle méthode sur un signal issu de l'analyse d'un mélange physique laisse douter de la légitimité du raisonnement.

Nous allons éclaircir certaines limitations de ces méthodes en prenant l'exemple de solutions contenant plusieurs molécules. Soit S_1 une première solution contenant une molécule A à la concentration a_1 et une molécule B à la concentration b_1 . Soit S_2 une deuxième solution contenant la même molécule B à la concentration b_2 et une molécule C à la concentration c_2 . Nous pouvons mélanger ces deux solutions pour en obtenir une troisième notée S_m de la manière suivante :

$$S_m = \alpha S_1 + (1 - \alpha) S_2 \quad (2.1)$$

où α est un coefficient de mélange compris entre 0 et 1.

De cette manière la solution S_m apparait clairement comme un mélange de deux sources S_1 et S_2 , mais peut aussi être décrite comme un mélange des trois molécules A, B et C. Si l'ambiguïté sur les sources à considérer n'est pas dérangeante dans le problème direct, le problème inverse pâtira grandement en l'absence de connaissances sur les sources initiales. Prenons l'exemple d'une autre solution S_p contenant les molécules A, B et C, mais dans des proportions telles qu'il n'existe pas de α tel qu'elle ne puisse pas s'écrire comme une combinaison de S_1 et S_2 . Un exemple évident de ce cas serait lorsqu'une des molécules considérées se retrouve à une concentration plus faible que le minimum dans S_1 ou S_2 , ou inversement lorsque la concentration dépasse le maximum. Cette solution S_p ne peut alors s'écrire que comme une combinaison des trois molécules considérées. Dans un cas réel de déconvolution, ni le nombre de sources ni leur complexité ne sont connus à l'avance.

Supposons que cette ambiguïté est levée et considérons maintenant l'étude des chromatogrammes du point de vue de la convolution et de la déconvolution. Un chromatogramme est un signal à deux dimensions issu de l'analyse d'une solution par une chromatographie couplée à la spectrométrie de masse. Nous désignerons par \mathbf{Y} un chromatogramme quelconque constitué des termes $y_{m,t}$, $m = 1..M$, $t = 1..T$. Un chromatogramme contenant S molécules de chromatogramme \mathbf{x} peut être décomposé en plusieurs composantes :

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.2)$$

où \mathbf{W} est une matrice de poids des composés présents dans le signal, \mathbf{X} est la matrice des signaux sources, $\boldsymbol{\beta}$ le signal de la ligne de base liée à la méthode utilisée et $\boldsymbol{\epsilon}$ du bruit. Les matrices de ligne de base et de bruit prennent la forme du chromatogramme, \mathbf{W} et \mathbf{X} sont décrites ci-dessous :

$$\mathbf{W} = (w_1 \mathbf{I}_M \quad w_2 \mathbf{I}_M \quad \dots \quad w_S \mathbf{I}_M)$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_S \end{pmatrix}$$

Le problème inverse consiste à retrouver \mathbf{X} et \mathbf{W} . Plusieurs problèmes émergent en tentant d'isoler la matrice des sources :

$$(\mathbf{W}'\mathbf{W})\mathbf{X} = \mathbf{W}'(\mathbf{Y} - \boldsymbol{\beta} - \boldsymbol{\epsilon}) \quad (2.3)$$

Nous ne connaissons que \mathbf{Y} donc il reste un grand nombre d'inconnues (\mathbf{W} , $\boldsymbol{\beta}$ et $\boldsymbol{\epsilon}$), la taille de \mathbf{W} n'est pas connue, il y a au maximum autant de sources que de molécules dans la solution analysée, et la matrice $\mathbf{W}'\mathbf{W}$ n'est pas inversible. Il est alors nécessaire de déterminer son pseudo-inverse $(\mathbf{W}'\mathbf{W})^+$ au sens de Moore-Penrose. Nous obtenons alors l'équation suivante :

$$\mathbf{X} = (\mathbf{W}'\mathbf{W})^+ \mathbf{W}'(\mathbf{Y} - \boldsymbol{\beta} - \boldsymbol{\epsilon}) \quad (2.4)$$

2.3.2.2 . Les techniques utilisées en chimiométrie

Il existe de nombreuses techniques de chimiométrie s'appuyant sur les statistiques des données fournies, le Tableau 2.6 présente trois méthodes courantes.

TABLEAU 2.6 – *Techniques communes utilisées en chimiométrie*

Nom	But	Contraintes
PCA (analyse en composantes principales)	Projection des données suivant leurs principaux axes d'inertie	Les composantes ou facteurs identifiés doivent pouvoir correspondre à des entités ou à des effets physiquement séparables.
ICA (analyse en composantes indépendantes)	Recherche de signaux sources indépendants	Méthode entièrement pilotée par les données (statistique d'ordre élevée). De grandes séries sont nécessaires pour exploiter la non-gaussianité des composantes. Les signaux sources ne sont pas nécessairement physiques (positifs).
PCoA (analyse en coordonnées principales)	Reconstruction d'une topologie (espace de coordonnées) probable à partir des distances paire à paire notamment pour la réalisation de classement	Capacité à mesurer des distances paire à paire.

L'analyse en composantes indépendantes a été développée dans les années 80 afin de répondre à des problèmes de séparation de sources à l'aveugle rencontrés dans la détection radar, les orchestres et de la forme suivante :

$$y = Mx + v \quad (2.5)$$

où x contient les signaux indépendants avant mélange aussi appelés sources, y les signaux mélangés dans différentes proportions, v le bruit et M est la matrice de mélange devant contenir au maximum autant de colonnes que de lignes, c'est à dire avoir plus de réalisations de signaux mélangés que de sources indépendantes.

Pour décrire l'idée derrière l'analyse en composantes indépendantes, nous allons considérer une analogie avec un orchestre de musique. Un orchestre jouant une partition donnée est composé de différents corps d'instruments de tailles variables, l'ensemble des musiciens d'un même corps correspondent à une seule source indépendante, le nombre d'instruments n'influant que sur l'intensité des notes jouées. La matrice de mélange dans cas est alors une fonction de la partition et du nombre de musiciens présents à une répétition donnée. Une personne à l'ouïe suffisamment entraînée sera alors capable à l'écoute de séparer les contributions de chacun des corps et d'estimer leur contribution à chaque répétition.

Une étude préliminaire de l'applicabilité de plusieurs techniques de chimiométrie sur des données de chromatographie en phase gaz par espace de tête couplée à la spectrométrie de masse (HS-GC-MS) a été réalisé par Kassouf et al. (2013). Dans cet article les auteurs appliquent à tour de rôle l'analyse en composantes principales (PCA), l'analyse en composantes indépendantes (ICA) et l'analyse en composantes communes et poids spécifiques (CCSWA) pour tenter de discriminer des six lots de granulés, préformes et bouteilles en PET issus de deux résines différentes. Un premier constat est que les trois méthodes utilisées ne conduisent pas à des résultats similaires d'après les auteurs qui ont eu recours à une évaluation subjective des données de sortie pour montrer des différences entre les échantillons en terme de présence ou d'absence de molécules, aucun critère objectif n'est donné pour classer les trois méthodes en terme de performances.

Nous pouvons reprocher aussi le plan d'échantillonnage, l'intention de discriminer l'influence du granulé utilisé et des étapes de mise en forme est louable, mais tel que présenté il n'y aucune certitude sur les liens de causalité entre les granulés, les préformes et les bouteilles étudiés. En effet rien n'indique qu'un même lot de granulés ai servi à la fabrication des préformes, ni que ces mêmes préformes aient été soufflés pour donner les échantillons de bouteilles. Outre le procédé de fabrication, les échantillons n'ont donc pas les mêmes antécédents en terme de stockage et donc de potentielle contamination. Ces travaux ont été réalisés sur un nombre restreint d'échantillons (2 granulés, 2 préformes et 2 bouteilles, chacun avec trois répétitions), les techniques de chimiométrie utilisées ne peuvent alors fournir qu'un nombre inférieur de composantes, et certaines de ces composantes ont été subjectivement éludées par les auteurs, car ne contenant pas d'informations "pertinentes". Cette quantité limitée de données ne leur a pas permis de valider leurs résultats en appliquant ces méthodes sur des jeux de données tronqués ou permutés.

Nous avons réalisé une classification sur la base d'une extraction des données des figures de l'étude, celle-ci est présentée en Figure 2.12. Les classements n'apportent que peu d'information sur la qualité des grades de granulés, préformes et bouteilles. Les échantillons sont presque tous séparés sans classification logique, à part pour la PCA où les bouteilles sont séparées des granulés et préformes dans un arbre symétrique. Enfin, les auteurs auteurs ont tenté d'identifier certains des composés responsables (2.7) des différences entre les échantillons sur les différentes méthodes en s'appuyant uniquement sur des comparaisons de spectres issus de bases de données commerciales, ces identifications n'ont en aucun cas été validées par des bases de données internes contenant des indices de rétention ou par des ajouts dosés des étalons.

TABLEAU 2.7 – Composés détectés selon les techniques chimiométriques utilisées par Kassouf et al. (2013)

Composé	PCA	ICA	CCSWA
2-methyl-1,3-dioxolane	PC1(+)	ND	CC6
ethylene glycol	PC1(+)	IC5(+)	ND
toluene	PC1(+)	ND	CC6
ethylbenzene	PC1(+)	ND	CC6
xylene isomers	PC1(+)	ND	CC6
heptanal	ND	IC2(+)	ND
nonanal	PC1(-)	IC2(+)	ND
decanal	PC1(-)	IC2(+)	ND
diethyl phthalate	PC1(-)	IC1(+) IC2(+)	ND
diisobutylphthalate	PC1(-)	ND	ND
2,4-bis(1,1-dimethyl)phenol	ND	ND	CC4
dichlorobenzene	ND	ND	CC6

PCi : Composé détecté dans la composante principale i ; ICi : Composé détecté dans la composante indépendante i ; CCi : Composé détecté dans la composante commune i ; (+/-) : Intensité positive/négative du composé chimique dans la composante considérée ; ND : Non détecté

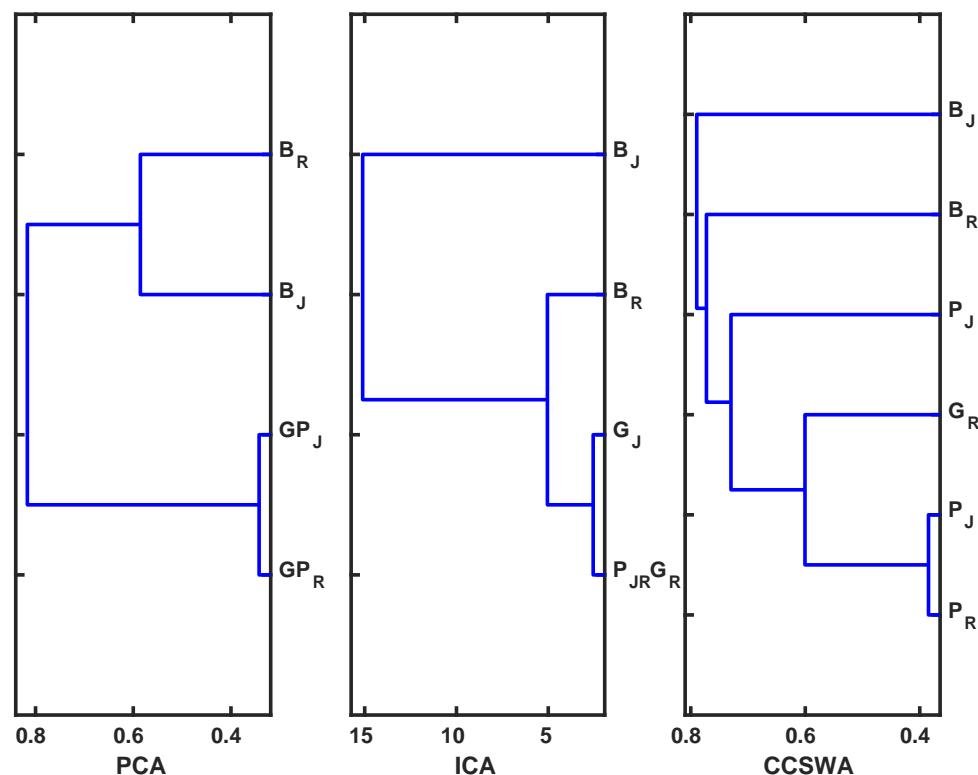


FIGURE 2.12 – Classification des échantillons d'après les trois méthodes utilisées par Kassouf et al. (2013). Les échantillons sont issus de deux grades de matériaux R et J, sous trois formes : granulés G, préformes P et bouteilles B.

2.3.2.3 . Les méthodes de poursuite

Le simple démixage des données n'est pas suffisant pour comprendre les effets et ou identifier les substances. Les sous-signaux indépendants peuvent ne pas être physiquement réalistes et de nombreuses substances peuvent présenter des structures similaires. Pour toutes ces raisons, il est préférable de rechercher une relation causale entre la présence probable d'une substance ou d'une combinaison de substances et les mesures réalisées. C'est l'hypothèse fondamentale utilisée dans la méthode de poursuite. Classiquement, la mesure est décomposée en deux fractions : les "causes" entraînant une forte corrélation et le résidu. Le résidu peut à son tour être redécomposé. [Nguyen et al. \(2015\)](#) a appliqué la méthode de poursuite à deux échelles en considérant à la fois l'introduction directe d'une substance ou d'une combinaison linéaire de substances. La méthode a été appliquée à la déconvolution de signaux de RMN ^1H en remarquant que les signaux provenant de sources physiques élémentaires peuvent être soit mesurés ou calculés par des théories moléculaires. L'approche vise à construire un modèle structurel où les mélanges sont reconstruits par des scénarios imbriqués et ramifiés en commençant par les substances les plus probables. Un algorithme de préférences généralisant les résultats de Condorcet sur les votes par préférence décide de l'ordre d'insertion des substances et des séquences. L'espace chimique d'un mélange n'est pas nécessairement connu parfaitement, mais la méthode génère à la fois des estimateurs parcimonieux des concentrations présentes (le signal est proportionnel au nombre de protons), mais également une empreinte du mélange dans l'espace des molécules insérées. Cette empreinte permet de reconnaître aisément tout mélange ou sous-mélange élémentaire similaire. Un mélange est codé sous la forme d'une séquence, d'un arbre de codes ou de lettres pouvant être comparé à une séquence similaire en acceptant des recombinaisons. Une modélisation structurale relativement similaire est également proposée par [Hyvärinen \(2013\)](#).

2.3.3 . Ce que la bioinformatique pourrait apporter

La bio-informatique fournit de nombreuses approches mathématiques pour traiter des données en volume avec des applications en génomique, protéomique et plus généralement dans les branches de biologie identifiées par le suffixe -ome ou -omique. Le suffixe est utilisé pour désigner une science intégrative combinant sur les données identifiées par le radical (prote-, lipid-, metabol-, transcript-) et le calcul. Comme pour les chatbots proposés comme démonstrateurs des nouvelles possibilités de l'intelligence artificielle, les approches intégratives ont gagné en audience avec la réussite de grands challenges comme le "human genome project" au début des années 2000 (Bayat, 2002). Dans le domaine des sciences analytiques appliquées aux aliments, le terme "foodomics" a été proposé (Cifuentes, 2009) pour appuyer les études en nutrition, la reconstruction de la traçabilité ou la recherche de fraudes (exemple de détection de la viande cheval dans les plats cuisinés) (Munekata et al., 2021) ou encore l'étude de molécules bioactives (Ortea, 2022).

Malgré les ressemblances, la bio-informatique ne doit pas être confondue avec la biologie computationnelle qui ajoute une dimension biophysique et biochimique à ses simulations (Navlakha and Bar-Joseph, 2011). L'hypothèse fondatrice est qu'il existe un code qui peut être compris soit comme le résultat d'un programme biologique ou informationnel ou comme le programme lui-même. L'idée de programme a été popularisée par l'ouvrage "The Selfish Gene" de l'ethnologue Richard Dawkins en 1976. Les développements se sont focalisés sur les techniques intensives qui permettent la reconnaissance de motifs, la fouille de données, l'auto-apprentissage et la visualisation. Le "human genome project" a stimulé le développement d'un langage de programmation spécifique le BioPerl construit comme une collection de modules Perl, un langage informatique dynamique développé par un linguiste pour traiter de manière efficace l'information de type textuel.

Les technologies de type ICA sont bien plus utilisées en métabolomique que pour identifier des contaminants. L'existence d'un codage de l'espace chimique des matériaux n'est pas établie. Le concept précédent de poursuite (Parie 2.3.2.3) esquisse toutefois la possibilité de coder les réponses des signaux mesurés à l'échelle de groupements chimiques (Gillet et al., 2011) ou des substances (Nguyen et al., 2015) et de les reconnaître très rapidement même quand ils sont mélangés à d'autres signaux. En utilisant l'analogie textuelle, les groupements chimiques seraient des syllabes et les substances des mots. Les séquences ordonnées de groupements peuvent être assemblés en mots, mais il n'existe pas de grammaire permettant d'assembler des mots en propositions, phrases et paragraphes qui traduirait une formulation, un usage, une origine géographique, un mode de collecte, de tri ou de décontamination d'un matériau. Les travaux précités considéraient toutes les substances en mélange sans tenir compte d'une éventuelle séparation physique (élation) qui permettrait de reconnaître séquentiellement les substances ou leurs groupements. Parce que les méthodes de déconvolution actuelles permettent de retrouver jusqu'à une dizaine de substances sans recourir à une séparation physique, une séquence probable pourrait être reconstruite même si les substances sont coéluées. Au pire, deux syllabes proches "MO" et "MAU" pourraient être confondues et vues comme l'une ou l'autre, ou leur mélange. Le texte réel ne serait peut-être pas complètement reconnu, mais il signerait pour un certain contexte. Les mêmes causes entraînant les mêmes effets, le contexte pourrait être reconnu. Par ailleurs, en associant des régions du texte peu ou pas identifiées, il serait quand même pos-

sible de reconnaître une portion de texte. Ces approches existent déjà et sont proposées pour la reconnaissance de plagiats (Foltýnek et al., 2019).

En informatique de nombreux outils ont été proposés pour comparer et aligner des textes de longueur variable ou pour reconstruire des différences entre deux fichiers. Dans le monde de Linux, l'algorithme "diff" est le plus populaire et est à l'origine des outils "rsync" et "git" qui permettent respectivement d'échanger et de stocker uniquement les opérations d'édition qui permettent de reconstruire le fichier modifié ou mis à jour à partir d'une copie du fichier initial. La longueur d'édition ou distance de Levenshtein (Levenshtein, 1965) est donnée par :

$$d_{i,j} = \min \begin{bmatrix} d_{i-1,j} + C_{del}(b_i) \\ d_{i,j-1} + C_{ins}(a_j) \\ d_{i-1,j-1} + (a_j \neq b_i) \cdot C_{sub}(a_j, b_i) \end{bmatrix} \quad (2.6)$$

où C_{del} , C_{ins} , C_{sub} sont les coûts fixés à l'avance d'une déletion, insertion et substitution.

L'Équation 2.6 permet de reconstruire des distances paire à paire entre des portions de texte. Les textes a_j et b_i sont comparés de manière matricielle pour tous les décalages possibles. Un exemple de matrice $d_{i,j}$ est donné pour la comparaison des mots "niche" et "chiens" dans le Tableau 2.8. Les déplacements en diagonale décrivent la substitution d'une lettre du mot "niche" par une lettre du mot "chiens", un déplacement vers le bas une suppression d'une lettre du mot "niche" et un déplacement vers la droite l'insertion d'une lettre du mot "chiens". Les meilleurs chemins étant ceux cumulant le score le plus bas possible.

Les termes de la diagonale donnent le coût d'une substitution, les termes au-dessus de la diagonale donnent le coût d'une déletion, et les termes inférieurs le coût d'une insertion. En réduisant les longueurs de décalage maximales et en translatant les résultats précédents, il est possible d'obtenir des implémentations rapides du calcul de la distance d'édition.

TABLEAU 2.8 – Exemple d'alignement des mots *niche* et *chiens* issu de Wikipedia (https://fr.wikipedia.org/wiki/Distance_de_Levenshtein, extrait le 10 janvier 2023)

		C	H	I	E	N	S
	0	1	2	3	4	5	6
N	1	1	2	3	4	4	5
I	2	2	2	2	3	4	5
C	3	2	3	3	3	4	5
H	4	3	2	3	4	4	5
E	5	4	3	3	3	4	5

Le chemin sélectionné est coloré. Les suppressions correspondent aux cases en orange, les insertions en vert et les substitutions (lettres identiques ou différentes) en gris

2.3.4 . Ce que la théorie de l'information pourrait apporter

La théorie de l'information couvre de nombreux aspects allant du codage, au stockage, à la transmission et à la quantification de l'information. C'est cette dernière caractéristique qui est la plus intéressante, car elle s'apparente à la notion de complexité. Certaines techniques de déconvolution utilisent indirectement des résultats issus de la théorie de l'information : principe de parcimonie ou minimisation de l'information mutuelle (ou mesure d'indépendance). Elles introduisent toutefois ces concepts a priori et on pourrait penser qu'elles auraient tendance à minimiser la présence de contaminants si ceux-ci sont présentés dans des proportions proches de limites de détection ou de linéarité des approches proposées. L'entropie est un des piliers de la physique théorique avec la mécanique classique, l'électromagnétisme de Maxwell, la relativité, la mécanique quantique et finalement la mécanique statistique de Boltzmann-Gibbs ([Tsallis, 2019](#)). Le prix Nobel de physique 2020 Roger Penrose, qui a été récompensé pour sa compréhension des objets singuliers que sont les trous noirs pense qu'ils laissent des traces dans le fond cosmologique même après leur évaporation ([An et al., 2020](#)). L'information ne disparaîtrait pas complètement même dans un trou noir du fait du rayonnement de Hawking. Les premières annonces spectaculaires de la détection d'un univers antérieur au nôtre et qui font suite à une première analyse des observations du détecteur à 70 GHz du satellite de Planck ont été révisées à la baisse ([Jow and Scott, 2020](#)), mais elles montrent comment la théorie de l'information pourrait être utilisée pour analyser objectivement des empreintes diffuses qui résulteraient de contaminations passées.

L'entropie initialement proposée Boltzmann-Gibbs est appelée entropie de Shannon dans le cadre de la théorie de l'information :

$$H = -k_B \sum_{i=1}^W p_i \ln p_i \text{ avec } \sum_{i=1}^W p_i = 1 \quad (2.7)$$

où k_B est la constante de Boltzmann, W le nombre de possibilités microscopiques et p_i leur probabilité.

Le frère de Roger Penrose, Oliver Penrose, a été le premier à formulé que l'entropie totale de deux sous-systèmes indépendants (notés ici L et C), $H(L+C)$, devaient être additives (([Penrose, 1970](#)) p167). La démonstration du caractère extensif de l'entropie est triviale, en remarquant que l'indépendance des sous-systèmes garantit que $p_{i,j}^{S+C} = p_{i,j}^S + p_{i,j}^C$, nous avons :

$$H(L+C) = H(L) + H(C) \quad (2.8)$$

Si L désigne la contamination issue d'une économie linéaire et C celle liée à une économie circulaire. L'entropie informationnelle combinera les deux informations. La différence $H(L+C) - H(L)$ caractérise la seule contribution liée à l'économie circulaire.

2.4 . Conclusions

L'économie circulaire complexifie le problème de gestion de la migration. On passe d'une liste de substances déclarées et/ou à contrôler (cas des plastiques, papiers et cartons en France, encres...) à une liste ouverte de substances. Le recyclage mécanique avec ou sans solvant propose de réutiliser le matériau sans revenir à une brique moléculaire élémentaire qui assurerait le retour à une pureté originelle. Les technologies tout comme l'origine des matériaux sont très disparates. Le changement de couleur des bouteilles de PET recyclées vers le gris du fait de mélange de gisements de différents PET colorés témoigne de la contamination résiduelle de la matière. Les azurants initialement ajoutés (IAS) sont encore présents dans la matière recyclée, où ils deviennent des NIAS. La littérature scientifique tant dans le domaine de la chimie analytique, emballage ou de l'environnement tend à mettre en avant les bénéfices du recyclage du point de vue du premier principe de la thermodynamique : moins de travail et d'énergie sont nécessaires pour produire un nouveau matériau. La question NIAS est généralement traitée en prenant le point de vue des composés de dégradation des substances initiales. Ce chapitre montre que la circularité comme tout cycle thermodynamique obéit également au second principe de la thermodynamique : l'entropie du système emballage + aliment et son milieu anthropisé augmentera inexorablement. Il ne s'agit pas ici d'invoquer la perte de chaleur (une énergie dégradée perdue pour l'homme), mais la dispersion des substances. Les étapes de décontamination peuvent réduire les quantités, mais elles ne peuvent pas réduire le nombre d'espèces.

L'augmentation inexorable du nombre de contaminants ou d'impuretés (terme au choix) pose une très grande difficulté à la chimie analytique en tant que discipline et à la chaîne de valeur du fait des conséquences d'une mauvaise décision. Aujourd'hui nous n'avons pas assez de recul pour identifier les étapes les plus critiques et où devraient être placés les points de contrôle. L'un des brouillons du nouveau règlement (UE) 2022/1616 proposait un contrôle de la matière recyclée toutes les vingt tonnes. La proposition n'a pas été retenue, car jugée trop contraignante, non seulement par sa fréquence, mais aussi par l'impossibilité physique de constituer un mélange représentatif à partir de solides divisés sur une masse aussi grande. Qu'elle est la valeur d'un échantillonnage sur des impuretés ? Mélanger des échantillons risque de les diluer au point qu'elles ne soient plus détectables. On pourrait argumenter que les méthodes sophistiquées de spectroscopie de masse haute-résolution et de chimiométrie décrites dans la littérature seraient capables d'identifier voire quantifier des substances fortuites dans des extraits de polymères. Ces travaux sont intéressants, mais ils n'abordent pas les points les plus cruciaux : la capacité à être utilisée en routine à des coûts qui ne renchérissent pas le prix des emballages pour le consommateur, la capacité à fonctionner à haut débit, la capacité à fournir une aide à la décision objective dans toute situation (matériau, recycleur, transformateur, utilisateur). Ce dernier point est d'autant plus important que le coût d'un rappel de produit alimentaire est sans commune mesure avec le coût de la matière première et que les décisions de type de gestion de risque pour les substances les plus toxiques qui doivent être prises pour des concentrations dans les matières recyclées qui sont inférieures aux limites de détection.

Ce diagnostic est confirmé par trois programmes de recherche collaboratifs financés par l'Agence Nationale de la Recherche et auxquels participent le LNE : PolySafe (ANR-20-CE34-0009), FoodSafeBioPack (ANR-20-CE21-0009) et PackSafe (ANR-21-CE21-0004). Le nombre croissant de substances détectées avec ou sans étape de reconcentration dans les échantillons étudiés traduit bien une réalité du terrain du fait du réemploi, du recyclage et/ou des contaminations croisées. Le LNE en tant que laboratoire national de référence doit pouvoir fournir des résultats opposables dans le cadre de lois existantes ou de bonnes pratiques entre industriels. Pour toute mesure et son interprétation, qu'elle que soit sa nature, deux risques doivent être considérés :

- Le risque de première espèce : le danger ou la non-conformité n'a pas été détectée (risque pour l'homme)
- Le risque de seconde espèce : la matière recyclée est déclassée à tort (augmentation des impacts pour l'environnement liées à une plus grande utilisation de matière vierge, risque de non-conformité par rapport aux obligations environnementales).

Ces deux risques sont rarement discutés dans les publications scientifiques s'intéressant aux NIAS. Le concept de validation croisée effleure une partie du problème, mais ne traite pas des comparaisons entre laboratoires et des effets de l'échantillonnage ou du type de mesure. Est-ce que l'on peut accepter un risque de 1% pour la mise sur le marché de matière recyclée impropre ? En France, chaque français produit 580 kg/an d'ordures ménagères (données ADEME publié en 2022 ([ADEME, 2022](#)), dont la moitié au moins serait des emballages alimentaires. En considérant que le volume d'aliment emballé est 50 fois plus important que l'emballage, un risque de 1% conduirait chaque français à ingérer 145 kg/an (0.39 kg/jour) d'aliments contaminés. Cette analyse explique notamment la crise associée à la contamination des aliments secs par les papiers et cartons recyclés : pas de contrôle, pas d'étape de décontamination.

Il ne faut pas s'y tromper. Les réglementations actuelles autorisent déjà un droit pour les industriels à contaminer les aliments et donc à exposer les consommateurs. L'inertie n'a pas besoin d'être totale (infaisable), mais contrôlable. Le Tableau 2.9 compare les décisions alternatives pour IAS et NIAS du point de vue d'un laboratoire de contrôle. Il n'est peut-être pas opportun de parler de révolution environnementale, mais il s'agit bien d'une révolution industrielle. Durant la période de transition, l'innovation doit être stimulée et encouragée tout au long de la chaîne de valeur : production, transformateurs et recyclage de matériaux, chaîne de transformation et de conditionnement des aliments, circuit de distribution et de collecte des matériaux, outils de gestion de la sécurité sanitaire. Le risque IAS est déjà bien géré au niveau des laboratoires de référence, notamment dans le cadre du réseau européen géré par le Centre de recherche Commun de la Commission Européenne. Les NIAS fortuits restent un objet de recherche plus hypothétique que pratique. Les NIAS ubiquitaires font par contre peser des risques plus importants et ne font pas l'objet de méthodes standardisées : pas de méthodes haut-débit, pas de bases de données de référence, pas de capacité à comparer la sécurité d'un emballage actuel avec celle d'un emballage du passé.

Cette partie bibliographique confirme la nécessité d'explorer des outils et les méthodes de laboratoire bien au-delà de celles déjà utilisées par la filière emballage, industrie alimentaire et

TABLEAU 2.9 – *Tableau récapitulatif de types de contaminants et leurs risques sanitaires associés*

Type de contaminant	IAS (liste positive ou soumise à contrôle)	NIAS fortuite	NIAS ubiquitaire
Fréquence	Systématique	Rare	Systématique
Echantillonnage	Aisé	Très difficile	Possible
Identification	Aisée avec étalons	Difficile	Difficile mais possible si récurrent
Détection	Aisée	Possible si concentration élevée	Possible si concentration élevée
Quantification	Possible avec gamme étalon	Très difficile	Semi-quantification au mieux
Applicable en routine	Déployé dans une majorité de laboratoires	Réservé aux laboratoires de recherche	Réservé aux laboratoires de recherche
Conséquence immédiate du risque de première espèce	Non-conformité	Aucune à rappel de produit si identifié après mise sur le marché	Variable selon que le danger est connu au préalable ou non
Chronicité du risque de première espèce	Faible	Faible	Forte
Conséquence immédiate du risque de seconde espèce	Changement de fournisseur et/ou de matière première	Changement de fournisseur et/ou de matière première	Augmentation signification de l'impact environnement

industrie du recyclage. La charge de la preuve de la salubrité des aliments emballés se déplace de plus en plus de l'aval de la filière vers l'amont. Parce que tout n'est pas identifiable et quantifiable, la traçabilité et les étapes pouvant conduire à une dispersion des contaminants (mauvais tri, lavage avec des résidus de produits) doivent être mieux contrôlées. Il n'existe pas de méthode actuellement pour démontrer l'origine d'un matériau et le respect des procédures. L'appel à manifestation d'intérêt 2022-2023 de l'Agence Nationale de la Recherche "Emballages et contenants alimentaires et politiques de durabilité : nouvelles contraintes d'alimentarité et impacts sur la nutrition et la santé" ([ANR, 2022](#)) témoigne d'immenses besoins en recherche. La possibilité de faire une analogie entre information chimique, entropie et mécanismes de transferts offre une opportunité de recherche pour des travaux exploratoires. Il y a toutefois un grand pas à réaliser entre mesure chimiométrique et code chimique réutilisable et opposable.

3 - Objectifs et démarche

3.1 . Objectifs généraux

De l'aveu même de l'Agence Nationale de la Recherche (ANR) dans son appel à manifestation d'intérêt (AMI) daté du 19 Septembre 2022, l'économie circulaire appliquée aux matériaux d'emballages affecterait de manière significative "la possibilité pour les acteurs des filières agroalimentaires de fournir collectivement des garanties sur l'innocuité des emballages". L'AMI identifie les évolutions législatives et réglementaires comme cause principale des nouveaux risques "en conduisant à des substitutions de matières [y compris de matières recyclées], des allègements [de matière], du réemploi.".

Les travaux présentés s'inscrivent dans la même perspective d'un risque accru de contamination des gisements, des matériaux au contact des aliments et surtout une diversification des contaminants et des voies de contamination. Il s'agit principalement d'explorer un nouveau paradigme qui est déjà partiellement utilisé dans la notion de migration globale du règlement (UE) plastique n° 10/2011 : la "contamination" de l'aliment (aussi appelée migration dans l'aliment) est l'ensemble des substances transférées avec une interprétation massique et non-toxicologique. Par exemple, tout matériau ne doit pas relarguer plus de $10 \text{ mg} \cdot \text{dm}^2$ d'un contaminant au contact de l'aliment. Cette interprétation est toutefois insuffisante pour identifier un danger potentiel à partir d'une molécule inconnue très fortement毒ique. Il faudrait disposer non seulement des quantités transférées mais aussi du profil toxicologique de la substance. Parce que beaucoup de substances non-intentionnellement ajoutées (NIAS) n'ont pas été évaluées ou de manière incomplète, d'autres concepts sont aujourd'hui explorés comme le principe d'une sécurité chimique équivalente. Ce principe exploré dans l'UMT SAFEMAT 22.07 entre le LNE (541 DEC Département Propriétés Chimiques des Matériaux) et AgroParisTech/INRAE (UMR 0782 SayFood) peut être vu soit pour une substance, soit pour une famille de substances. Pour une substance donnée, ce principe permet de justifier qu'un matériau "contaminé" avec une barrière fonctionnelle peut être aussi sûr qu'un matériau "décontaminé" tant que la barrière est capable de retarder suffisamment les transferts. Pour une famille de substances présentes dans un matériau inconnu (ex. issu de l'économie circulaire), nous pourrions penser que la sécurité sanitaire serait équivalente tant que le profil des contaminants resterait similaire à celui rencontré dans un matériau ou gisement considéré acceptable. A titre d'illustration, la présence de contenants non alimentaires (cosmétiques, détergents, peinture) ou ayant subit de mauvais usages (stockage d'herbicide, fluides mécaniques) affecte le profil des substances tout en laissant inchangés les autres paramètres (nature du matériau, aspect...). Les travaux visent à poser les bases d'un concept d'empreinte chimique qui permettrait par soustraction mathématique d'extraire les modifications du profil des contaminants potentiels, identifiés arbitrairement comme toute substance volatile ou semi-volatile présente dans le matériau. Le matériau peut être de toute nature : un thermoplastique, un thermodurcissable, un matériau naturel (papier, cartons), un matériau vierge ou recyclé, un produit fini, un gisement.

Parce qu'il y a peu d'antériorité sur ces notions, la thèse prend le parti pris que la notion d'empreintes est à rapprocher du concept d'information "chimique" et aurait une interprétation dans le cadre de la thermodynamique des phénomènes irréversibles. L'entropie un terme inventé par Rudolf Clausius est également une mesure de l'information contenue dans un signal, notée H . La phrase A ="le contaminant est transféré" contient moins d'information que la phrase est B ="le contaminant est néoformé" L'équivalence existe de manière formelle entre information et contaminants si l'empreinte associe des molécules entières (des mots) et non leurs fragments (lettres ou syllabes). Le signal contient également une partie de l'information qui n'est pas associée aux processus physiques (transferts, mélanges de gisements, actes malveillants) et chimiques (réactions) qui contrôlent la contamination des matériaux. La présence d'information non codante complique toutefois la construction d'une échelle chimique pour une méthode analytique ou spectroscopique, ainsi on l'inégalité $H(A) < H(B)$ mais on a $H(transfr) > H(noform)$. Ces travaux visent à poser les bases de la construction d'une échelle de l'information chimique qui mesure l'information ajoutée ou enlevée qui est présente dans l'échantillon et qui doit être indépendante du processus de mesure (extraction, analyse). D'autres complications peuvent affecter l'existence une échelle commune entre matériaux, gisements et pratiques capables d'estimer le nombre et le type de contaminants (notre ignorance) et sa monotonie théorique qui mesure l'irréversibilité transferts et des réactions. L'information instrumentale comprend deux parties, une mesure en intensité (proportionnelle aux quantités) et une mesure qualitative associée aux fonctions chimiques et/ou fragments présents. L'empreinte issue de la mesure de l'information ne sera pas nécessairement linéaire avec le nombre et la variété de substances. De la même manière, le dénombrement des substances dans un gisement ou un matériau ne préjuge en rien de leur origine, de la causalité des transferts et des réactions. Et le transfert d'une empreinte ou d'informations ne traduit pas la sélectivité des transferts (effet de taille et d'affinité) et des réactions (oxydation/scission, hydrolyse/transestérification).

La thèse explore la possibilité de traduire des signaux de chimie analytique en "information chimique", traitable avec des concepts d'extensivité qui permettent de réaliser des opérations algébriques (soustractions) pour séparer les causes probables, dont l'indépendance statistique est garantie par construction expérimentale ou parce qu'elle est jugée suffisamment probable. Dans ce cadre, l'information est transférée si deux matériaux ou gisements sont mélangés (transferts complets avec travail mécanique) ou si deux matériaux ou un matériau et un aliment échangent des substances (transferts spontanés mais partiels). Les réactions chimiques obéissent aux mêmes principes en comparant un matériau à un état du même matériau dans le passé.

Il s'agit ainsi : (i) d'en illustrer les principes et d'en poser les bases pour une utilisation plus générale dans le cadre de l'économie circulaire :

- extraction d'un code irréductible qui signerait un matériau, sa formulation, son état après transformation/collecte/décontamination
- codification normalisée pour plusieurs échantillons et applicable à plusieurs techniques
- identification rapide d'empreintes de NIAS, de contaminations ou réactions fortuites

(ii) d'en identifier les difficultés de mise en œuvre :

- invariance ou non avec l'opérateur et l'équipement utilisé
- analyse en nombre de composés plutôt qu'une analyse en quantité de composés transférables
- compatibilité avec les approches utilisées par le laboratoire national de référence (LNE) et les outils de chimiométrie existants
- compatibilité avec d'autres développements visant à automatiser l'analyse des risques (évaluation de la migration, exposition du consommateur) sur la base de classes de substances "probables" identifiées en chimie-analytique et d'une catégorisation semi-automatique de la classe toxicologique d'appartenance

3.2 . Objectifs spécifiques

Sans préjuger de la capacité à extrapoler l'approche à d'autres techniques analytiques (ex. spectroscopie vibratoire, résonance magnétique nucléaire), le concept d'information chimique est étudié pour des mesures en chromatographie en phase gaz couplée la spectrométrie de masse basse résolution en injection liquide (GC-MS) ou en espace de tête (HS-GC-MS). Il s'agit d'un parti justifié et assumé par la nécessité de déployer l'approche auprès d'un plus grand nombre acteurs (opérateurs privés, publics, autorités). Ce qui ne serait pas possible avec des techniques haute-résolution plus coûteuses et plus difficiles à interpréter.

La notion d'information chimique proposée dans les objectifs généraux permet de répondre à une grande partie des nouvelles exigences pour les matières plastiques recyclées tout en restant également applicable aux matériaux cellulosiques. La question globale d'information chimique est déclinée en objectifs spécifiques plus facilement accessibles :

- O1 : traitement robuste (transformation) des chromatogrammes de GC-MS vis à vis des biais de la méthode (saignement de colonne, bruit, déplacements) tout en conservant les propriétés du signal (corrélation des ions)
- possibilité d'inversion de la transformation pour une utilisation par l'opérateur via les outils conventionnels ou chimiométriques
- codage de l'information chimique sous la forme de mots ou phrases
- O2 : réduction du poids des chromatogrammes pour permettre un stockage sur le long terme
- O3 : comparaison de l'information chimique, mise en évidence les différences et similitudes
- O4 : application à des besoins actuels
 - suivi d'un procédé
 - suivi de gisement de matériaux recyclés
- O5 : développement d'un logiciel open source compatible à terme avec les outils d'évaluation du risque chimique

3.3 . Démarche

Chaque objectif spécifique identifié précédemment fait l'objet d'une démarche pour tenter d'y répondre avec des attentes précises en termes de résultats présentés dans la table 3.1. Certaines démarches seront ultérieurement développées lorsque qu'elles ont largement impacté les choix finaux de la méthodologie.

TABLEAU 3.1 – *Démarche mise en place pour répondre aux objectifs*

Objectif	Démarche	Résultats escomptés
O1	La méthode va s'appuyer sur des propriétés mathématiques générales du signal afin de ne pas le dénaturer et conserver les zones d'intérêt, être inversible ou pseudo-inversible pour assurer la compatibilité avec les outils traditionnels, et la transformation doit faciliter la recherche de pics et spectres	La méthode doit améliorer la répétabilité et reproductibilité de la GC-MS, une méthode d'inversion doit être proposée avec des limitations documentées, la recherche de pics et de spectres doit pouvoir se faire facilement avec des expressions régulières
O2	La méthode doit réduire le poids des données en combinant coarse-graining et compression multi-échelle	Le poids doit être au moins réduit de 50% sans perte de l'information générale
O3	La méthode va s'appuyer sur le concept d'alignement de séquences utilisé en génomique et sur des critères de distance objectifs tels que la distance de Levenshtein et la notion d'entropie en théorie de l'information	Comparaison paire à paire des chromatogrammes avec deux critères de distance permettant une cartographie des échantillons
O4	La méthode sera appliquée aux deux catégories de jeux de données, les résultats seront analysés en détail pour chercher des marqueurs de dégradation de procédé dans le premier cas ou de contamination dans le deuxième	Cartes des échantillons pour les deux applications et liste de composés marqueurs
O5	L'ensemble des codes utilisés seront mis à disposition avec une documentation exhaustive et seront ouverts à toute modification ou traduction dans un autre langage, un wrapper sera fourni pour faciliter l'utilisation aux utilisateurs novices	Toolbox MatLab complète avec documentation (nécessite d'avoir les dépendances), équivalent en pseudo-code pour l'implémentation dans d'autres langages, wrapper "presse bouton" pour une utilisation basique

3.4 . Théorie et choix du traitement du signal

3.4.1 . Choix concernant les propriétés spécifiques aux signaux rencontrés

Les signaux que l'on peut rencontrer en chimie analytique possèdent chacun des propriétés spécifiques qui sont traditionnellement utilisés dans le retraitement. Nous pouvons citer l'exemple de la chromatographie couplée à la spectrométrie de masse dont l'ensemble des ions corrélés à un instant signent pour une molécule particulière, ou encore la multiplicité des pics en RMN permettant de remonter à la structure d'un composé.

Bien que nous ne traiterons que le cas de chromatogrammes, nous avons fait le choix de développer une méthodologie générale devant être applicable à des signaux variés, pour cette raison nous ne nous appuierons pas sur ces spécificités, sans pour autant dénaturer le signal afin de pouvoir tout de même utiliser *a posteriori* ces propriétés qui font la force des techniques analytiques utilisées.

3.4.2 . Choix sur la transformée en ondelettes continue

La transformée en ondelettes analytique est un opérateur mathématique continu 2D (transformation affine) qui décompose un signal temporel $s(t)$ sur la base d'une fonction élémentaire continue appelée ondelette qui est translatée d'un pas τ et dilatée d'un facteur d'échelle ℓ . La projection sur la base d'ondelettes $\psi(\tau, \ell)$ est réalisée par un produit de convolution :

$$S(t, \ell) = \int_{-\infty}^{+\infty} s(t') \psi_{\ell, \tau}(t) dt \quad (3.1)$$

Le choix de la famille d'ondelettes se fait parmi les fonctions continues quarribles, de moyenne nulle et d'énergie unitaire. En notant Ψ , les ondelettes filles $\psi(\tau, \ell)$ générées par translation τ et dilatation ℓ sont données par :

$$\psi(\tau, \ell) = \frac{1}{\sqrt{\ell}} \Psi\left(\frac{\tau}{\ell}\right) \quad (3.2)$$

Parce que le produit de convolution de l'Équation 3.1 est un opérateur linéaire, les propriétés suivantes pourront être utilisées pour construire un modèle statistique des signaux chimiques :

Propriété 3.1 (Linéarité de la transformée en ondelettes) *La transformée à l'échelle ℓ d'une combinaison linéaire de signaux temporels $s_{i=1..n}(t)$ sous la forme $s_{tot} = \sum_{i=1}^n k_i s_i(t)$ est égale à la combinaison linéaire des signaux transformés $S_{i=1..n}(t, \ell)$ à l'échelle ℓ :*

$$S_{tot}(t, \ell) = \sum_{i=1}^n k_i S_i(t, \ell) \quad (3.3)$$

où les $k_{i=1..n}$ sont des poids scalaires.

La transformée en ondelettes permet une analyse en temps et fréquence (ou temps et période ℓ) sans avoir à recourir à des opérations de décimation ou de rééchantillonnage. En choisissant une échelle ℓ plus grande que la période d'acquisition d'un signal, les détails de haute fréquence sont perdus tout en conservant, voire amplifiant les basses fréquences.

Propriété 3.2 (Propriété de loupe continue sur signal transformé) En notant ζ le facteur de zoom, la transformée d'un signal $s(\zeta t)$ dilaté de ce facteur ζ est alors égale à ce même signal non dilaté $s(t)$ mais transformé par l'ondelette dilatée d'échelle $\zeta\ell$:

$$S(\zeta t, \ell) = S(t, \zeta\ell) \quad (3.4)$$

L'effet d'amplification est particulièrement intéressant pour les signaux faibles. Il est obtenu en remarquant la proximité entre les opérateurs de convolution (mesure de l'ondelette sur le signal) et de corrélation (mesure de la similarité entre l'ondelette et le signal). Les deux mesures sont équivalentes pour les ondelettes symétriques. Dans le cadre des signaux chromatographiques, les signaux sont généralement assimilés à des gaussiennes ou à des lorentziennes. Il apparaît donc naturel d'utiliser des ondelettes dérivant de l'expression d'une distribution gaussienne. Par ailleurs, l'opérateur de convolution permet de déplacer l'opérateur de dérivation du signal vers l'ondelette et inversement.

Propriété 3.3 (Distribution de l'opérateur de dérivation dans la transformé) La transformée d'un signal dérivé $\frac{\partial s}{\partial t}(t)$ par une ondelette $\psi(t, \ell)$ notée $\frac{\partial S}{\partial t}(t, \ell)$ est équivalente à la transformée du signal brut $s(t)$ par la dérivée de l'ondelette $\frac{\partial \psi}{\partial t}(t, \ell)$:

$$\frac{\partial S}{\partial t}(t, \ell) = \int_{-\infty}^{+\infty} s(t') \frac{\partial \psi}{\partial t}(t' - t, \ell) dt' \quad (3.5)$$

Les ondelettes mères issues de dérivées énièmes de gaussiennes (aussi appelée ondelettes hermitiennes) obéissent à la fonction génératrice :

$$\Psi_n(t) = (2n)^{-\frac{n}{2}} c_n H e_n(t) e^{-\frac{1}{2}t^2} \quad n \in \mathbb{N}^* \quad (3.6)$$

où $H e_n$ est le polynôme d'Hermite d'ordre n et c_n est le coefficient de normalisation d'ordre n donnée par : 3.7.

$$c_n = \left(n^{\frac{1}{2}-n} \Gamma \left(n + \frac{1}{2} \right) \right)^{-\frac{1}{2}} \quad n \in \mathbb{N} \quad (3.7)$$

où Γ est une extension de la fonction factorielle à l'ensemble des nombres complexes z tels que $Re(z) > 0$ définie par :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (3.8)$$

TABLEAU 3.2 – Propriétés des ondelettes hermitiennes d'ordres 1, 2 et 3 hermitiennes

Ordre de dérivation n	Lobes	Parité	Aire (par lobe)	Énergie (par lobe)
1	2	Impaire	0 (-0.5, 0.5)	1 (0.5, 0.5)
2	3	Paire	0 (-0.25, 0.5, -0.25)	1 (0.15, 0.7, 0.15)
3	4	Impaire	0 (-0.12, 0.38, -0.38, 0.12)	1 (0.04, 0.46, 0.46, 0.04)

Nous avons appliqué la transformée en ondelettes avec ces trois ondelettes sur un signal type de chromatographie présenté en Figure 3.2 constitué de bruit de grenade (suivant une loi de Poisson), d'une ligne de base avec une rampe et un plateau, trois gaussiennes identiques en aire

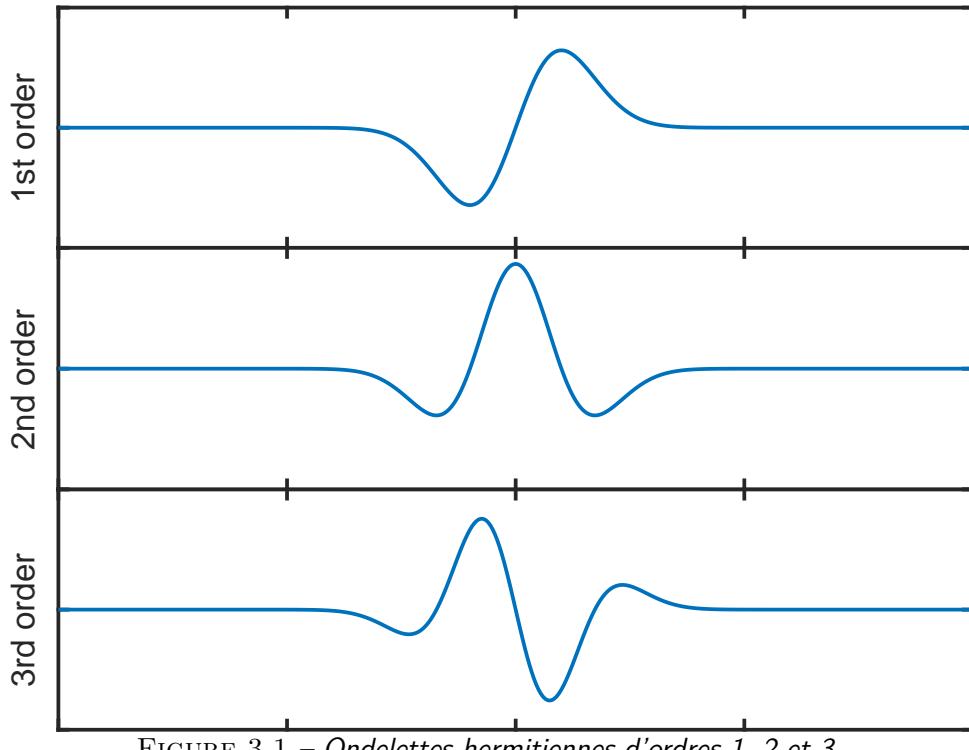


FIGURE 3.1 – Ondelettes hermitiennes d’ordres 1, 2 et 3

et en largeur disposées sur les trois zones d’intérêt. Nous pouvons voir que les trois ondelettes n’ont pas le même comportement vis-à-vis des différentes parties du signal, elles ont toutes filtré une partie du bruit ainsi que le plateau de la ligne de base, cependant l’ondelette d’ordre 1 est sensible à la pente de la rampe. Les ondelettes d’ordres 2 et 3 semblent être donc plus propices au nettoyage de chromatogrammes, mais nous ne conserverons que celle d’ordre 2, car elle conserve la position des maxima des gaussiennes et nous la désignerons par le nom d’ondelette de Ricker.

Si le choix de l’ondelette est crucial pour le traitement d’un signal, il ne faut pas non plus négliger l’influence de l’échelle ℓ sur le résultat de la transformée, il est même souvent nécessaire de faire une analyse multi-échelle. Nous avons appliqué la transformée en ondelettes avec l’ondelette de Ricker avec des échelles ℓ allant de 2^0 à 2^5 sur un signal constitué de bruit, d’une ligne de base et de gaussiennes d’écarts-types variables, le résultat est présenté en figure 3.3.

Nous pouvons observer que les signaux ne sont pas extraits de la même manière d’une échelle à une autre, la ligne de base est toujours filtrée, mais le bruit est majoritairement extrait aux basses échelles. Les gaussiennes sont extraites à des échelles plus hautes selon leur largeur, il existe donc une échelle optimale pour chaque gaussienne correspondant à 2.48 fois leur largeur tel que montré figure 3.4. Il nous paraît alors nécessaire d’appliquer systématiquement des analyses multi-échelle pour les signaux les plus complexes.

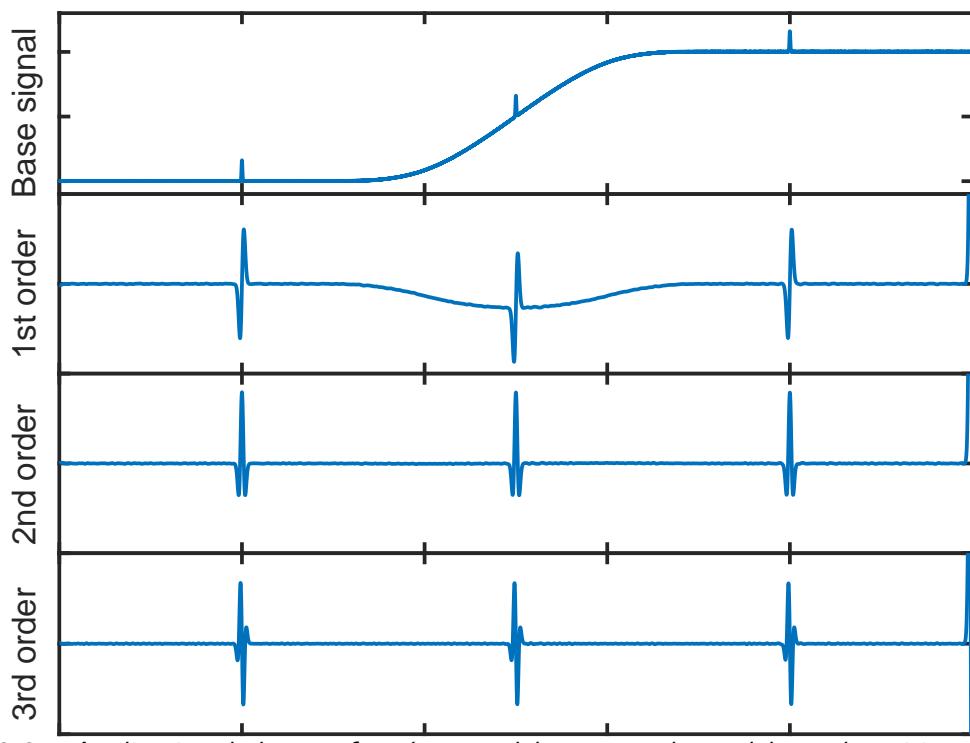


FIGURE 3.2 – Application de la transformée en ondelettes avec les ondelettes hermitiennes d'ordres 1, 2 et 3 sur un ersatz de signal chromatographique

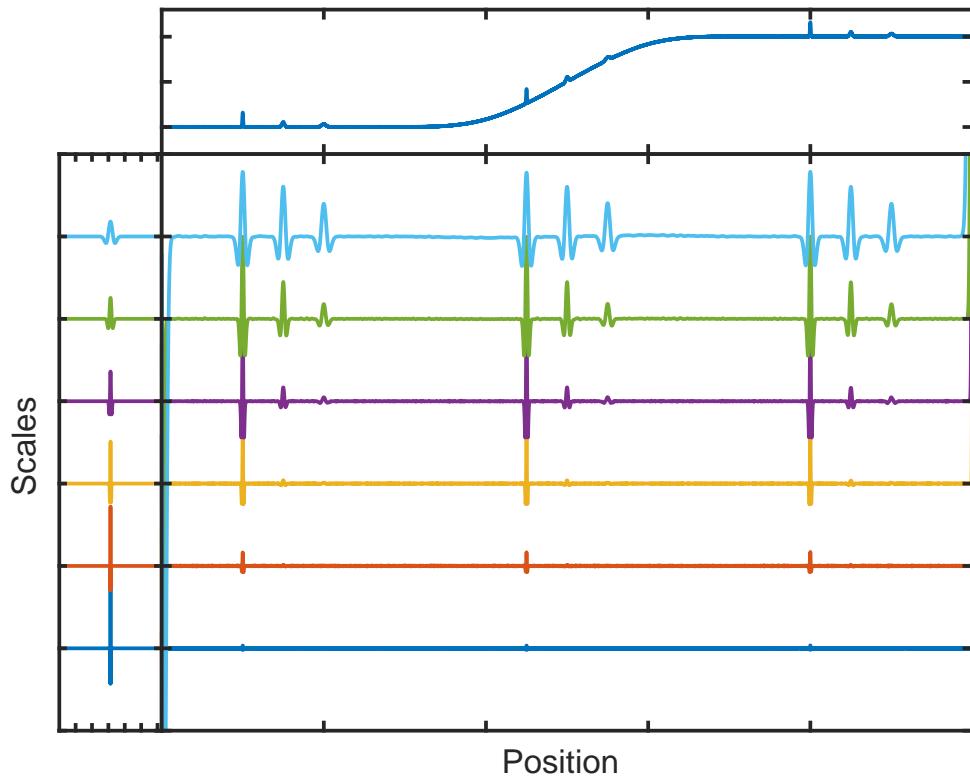


FIGURE 3.3 – Influence de l'échelle sur la transformée

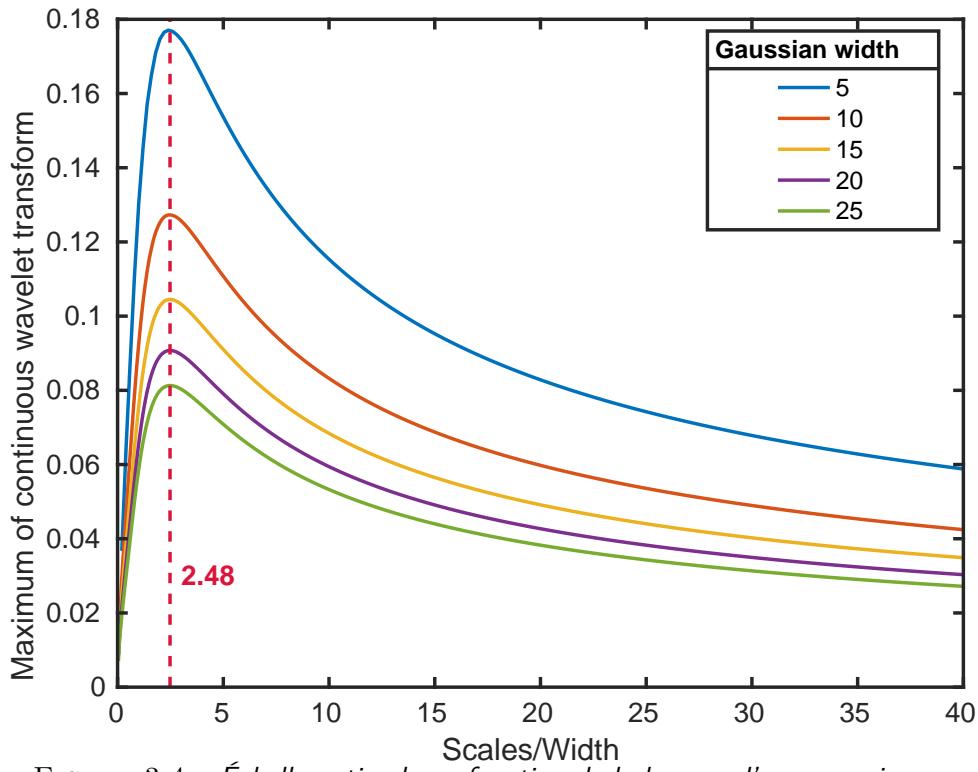


FIGURE 3.4 – Échelle optimale en fonction de la largeur d'une gaussienne

Un besoin de compatibilité avec les outils traditionnels de traitement de données et de chimiométrie a été remonté, ces outils fonctionnent tous dans l'espace initial des données, nous devons alors fournir un moyen de passer de l'espace de la transformée en ondelettes vers l'espace originel qui est un problème inverse mal posé. Des travaux ont été réalisés pour la détection de pics à l'aide de la transformée en ondelette (Du et al., 2006) dans lesquels sont posées les principes d'échelle optimale et de caractérisation de pics. Cependant les auteurs n'ont pas profité des capacités de filtrage de la transformée et de l'augmentation de sensibilité pour estimer l'aire des signaux d'intérêt.

3.4.3 . Choix sur l'encodage et la compression des données chromatographiques

Nous avons pu voir que la transformée en ondelettes continue avec l'ondelette de Ricker transformait les signaux chromatographiques strictement positifs en des signaux plus complexes avec des parties négatives. Nous avons développé dans les objectifs spécifiques et les démarches associées le besoin d'encodage du signal pour d'une part compresser le signal chromatographique en vue du stockage mais aussi d'autre part pour permettre l'utilisation des algorithmes d'alignement et de comparaison de séquences de caractères. Nous avons donc décidé d'utiliser les propriétés de monotonies pour décrire des segments entiers du signal à l'aide d'un nombre limité de caractères, réalisant en même temps une étape de coarse-graining et de compression permettant de réduire la quantité d'information nécessaire pour décrire le signal. Si nous ne considérions que le caractère croissant ou décroissant, chaque signal de serait qu'une alternance de ces deux segments et nous ne serions pas capables de différencier deux signaux. Nous profitons alors de la complexité apportée par la transformée en ondelettes pour introduire un encodage répartis selon trois paramètres :

- Monotonie croissante ou décroissante
- Début de segment positif ou négatif
- Changement de signe

Cela permet un encodage à six lettres présenté en figure 3.5 avec deux combinaisons de paramètres impossibles (croissant, début positif, croise l'axe des abscisses et décroissant, début négatif, croise l'axe des abscisses).

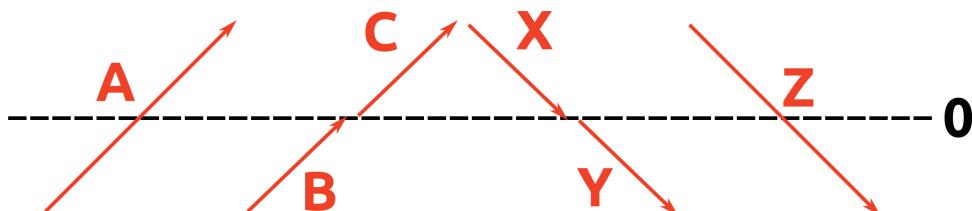


FIGURE 3.5 – *Encodage des six segments possibles.*

Aux lettres de l'encodage sont ajoutées la longueur et la hauteur du segment associé, permettant ainsi de décompresser les séquences de caractères de tous les ions d'un chromatogramme pour obtenir un texte rectangulaire. Il devient alors possible de réaliser une recherche rapide de pics ou ensemble de pics correspondant au spectre d'une molécule à l'aide d'expressions régulières.

La hauteur des segments sera utilisée pour filtrer ceux appartenant au bruit, pour cela nous allons utiliser plusieurs propriétés des ondelettes et de la transformée. Nous rappelons que l'ondelette de Ricker est normalisée pour avoir une énergie unitaire et que le bruit dans un chromatogramme est un bruit de grenaille modélisable par un processus de Poisson en chaque point et indépendants les uns des autres. La variance du signal colonne est ainsi conservée lors de la transformée.

La Figure 3.6 présente le niveau de compression donné par la transformée en de Ricker à une certaine échelle ℓ couplée à un encodage à six caractères. Le taux de compression prends en compte le nombre de segments pour décrire le chromatogramme ainsi que leur largeur et hauteur chacun encodé en type double (8 octets) également utilisé pour les données brutes. Nous pouvons observer que l'effet de compression ne s'opère qu'à partir des échelles $\ell \geq 1.85$, les données prennent même 50% plus de place à l'échelle $\ell = 1$. L'échelle ℓ optimale pour nos données typiques se situe à 8, à ce stade les données stockées sont cinq fois moins lourdes qu'initialement.

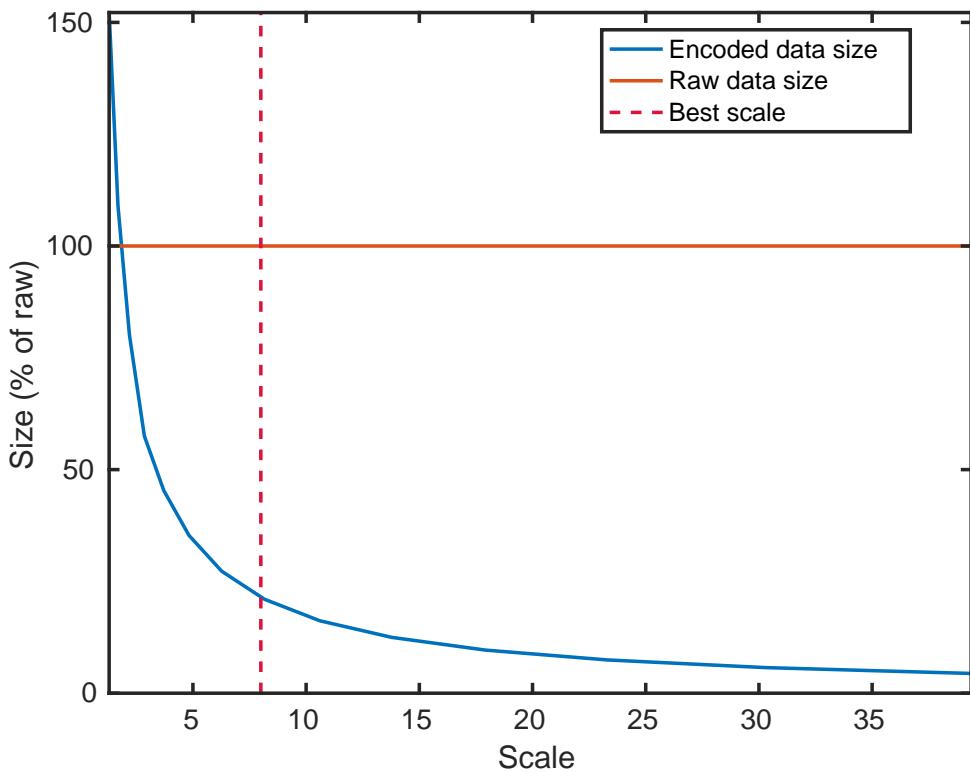


FIGURE 3.6 – Compression obtenue par l'encodage du signal transformé en fonction de l'échelle.

4 - Matériels et méthodes

4.1 . Matériels

Les échantillons de matière plastique utilisés pour reconstruire une carte des déplacements chimiques dans le cadre d'une économie linéaire et circulaire sont constitués de deux types d'échantillons :

- des échantillons prélevés sur un seul site industriel et soumis à des formulations et des conditions de mise forme contrôlées ;
- des échantillons collectés sur les marchés internationaux de la matière recyclée soit avant décontamination (chez les recycleurs), soit après décontamination (granulé commercial et proposé au contact alimentaire)

Les échantillons étaient disponibles sous la forme de granulés, de paillettes ou de produits finis. Deux matrices ont été retenues. Le polypropylène isotactique (PP) a été choisi comme modèle pour analyser l'effet du passage d'une économie linéaire à une économie circulaire sur l'espace chimique qui en résulte. Parce que le polyéthylène téréphthalate (PET) est largement recyclé à l'échelle globale, des échantillons de PET ont été collectés sur tous les continents (hors Afrique) de manière à construire une représentation de l'espace chimique représentative d'une économie circulaire généralisée et qui obéirait à des règles variables. Les procédés de collecte et de décontamination n'obéissent pas aux mêmes standards en fonction de l'endroit où ils sont recyclés. Le Tableau 4.1 donne une vue générale de tous les échantillons disponibles pour cette étude. Des échantillons de matériau vierge, non transformé (granulé) et de grade alimentaire sont inclus dans la liste de manière à établir des points de référence. Les échantillons ont été obtenus dans le cadre de projets de collaboration avec différents industriels et de la grande distribution. Les échantillons sont anonymisés de manière à permettre une analyse neutre des résultats. Seuls les effets qui ont pu être tracés (formulation, température, mode de tri ou collecte, mode de recyclage, origine) sont rapportés dans les résultats.

TABLEAU 4.1 – *Vue générale des échantillons utilisés pour construire les cartes des déplacements chimiques*

Catégorie	Cible	Nombre de sous groupes (nombre d'échantillons indépendants)	Origine	Nature
Échantillons de PP utilisés pour étudier l'effet de la formulation et des procédés de mise en forme à l'échelle d'un site de production unique	Économie linéaire	7 grades (23 échantillons)	France	Granulés et matériaux finis
Échantillons de PP	Économie circulaire	4 groupes (6 échantillons)	USA et Canada	Granulés recyclés et vierge
Échantillons de PET	Économie circulaire	25 groupes (136 échantillons)	USA, Canada, Amérique latine, Asie, Australie	Granulés et gisements, une bouteille d'eau de référence commerciale

4.1.1 . Échantillons de PP pour l'étude des effets de mise en forme et formulation

Le jeu de 23 échantillons en polypropylènes destiné au suivi d'un procédé est détaillé dans la table 4.2, pour chaque échantillon sont donnés leur état (granulés ou mis en forme), la référence de polypropylène, si le produit est coloré (noir de carbone ou dioxyde de titane) ou si une étiquette en PP est apposée par fusion (IML : In-Mold Labelling) et enfin à quelle température la mise en forme a été effectuée (220 °C en conditions normales, et 270 °C à 290 °C en conditions dégradées).

TABLEAU 4.2 – *Echantillons pour le suivi de procédé*

N° échantillon	État	Matériau	Coloration ou IML*	Température
P1	Granulés	Grade 1	/	/
P2	Granulés	Grade 2	/	/
P3	Granulés	Grade 3	/	/
P4	Granulés	Grade 4	/	/
P5	Granulés	Grade 5	/	/
P6	Granulés	Grade 6	Noir de carbone	/
P7	Granulés	Grade 7	Dioxyde de titane	/
P8	Barquette	Grade 3	Noir de carbone	220 °C
P9	Barquette	Grade 1	/	220 °C
P10	Barquette	Grade 1	IML	220 °C
P11	Barquette	Grade 4	/	220 °C
P12	Barquette	Grade 3	Noir de carbone	275 °C
P13	Barquette	Grade 1	/	270 °C
P14	Barquette	Grade 1	IML	270 °C
P15	Barquette	Grade 4	/	270 °C
P16	Couvercle	Grade 2	/	220 °C
P17	Couvercle	Grade 2	IML	220 °C
P18	Pot	Grade 1	Dioxyde de titane	220 °C
P19	Couvercle	Grade 2	/	290 °C
P20	Couvercle	Grade 2	IML	290 °C
P21	Pot	Grade 1	Dioxyde de titane	275 °C
P22	Barquette	Grade 4	Noir de carbone	220 °C
P23	Barquette	Grade 4	Noir de carbone	275 °C

*IML : In-Mould Labelling (étiquette fusionnée)

4.1.2 . Échantillons de PP et PET issus de l'économie circulaire

Les 124 échantillons utilisés pour le suivi et l'orientation de matériaux recyclés sont divisés en deux catégories suivant :

- qu'ils reproduisent de la variabilité de la matière recyclée et décontaminée destinée au contact alimentaire
- qu'ils permettent d'étudier l'origine du gisement de la première recyclée.

4.1.2.1 . Échantillons pour l'étude des origines de la matière recyclée

Une vue générale des différentes matières recyclées proposées au contact alimentaire est montré dans la Figure 4.1. Les échantillons incluent des granulés en polypropylène et en polyéthylène téréphthalate vierges ou recyclés d'origines. Ils sont a priori considérés conformes par rapport à leur zone géographique de recyclage. Ils ne sont pas nécessairement destinés à une exportation en dehors de leur région géographique de collecte.

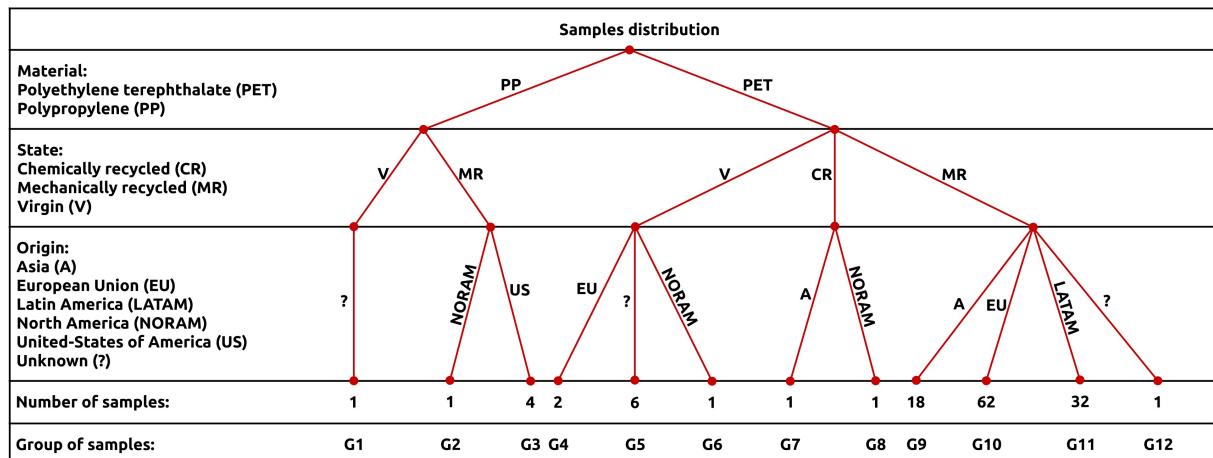


FIGURE 4.1 – Échantillons pour l'étude des origines de la matière recyclée. P : polypropylène ; PET : polyéthylène téréphtalate ; V : vierge ; MR : recyclé mécaniquement ; CR : recyclé chimiquement ; EU : Union européenne ; NORAM : Amérique de Nord ; LATAM : Amérique latine ; A : Asie ; US : États-Unis ; ? : Origine inconnue

4.1.2.2 . Échantillons pour l'étude de la traçabilité des gisements

Les 12 échantillons représentatifs des gisements sont détaillés dans la Tableau 4.3 contient des flocons de polyéthylène téréphtalate post-consommateur (i.e. récoltés après usage auprès des consommateurs puis broyés) non-décontaminés soit clairs(origine 100% alimentaire) soit colorés (contenant 5% de matériaux non-alimentaires).

TABLEAU 4.3 – Échantillons pour l'étude de la traçabilité des gisements

N° échantillon	Etat	Origine
G13-1 à G13-5	Paillettes	Mixte (5% non-alimentaire)
G14-1 à G14-5	Paillettes	Alimentaire
G15	Paillettes	Mixte de référence
G16	Bouteille	Contact alimentaire

4.2 . Méthodes

Les travaux s'appuient sur des empreintes chimiques obtenues expérimentalement après extraction (liquide ou gaz) suivi d'une analyse en chromatographie en phase gazeuse couplée à la spectrométrie de masse. L'empreinte numérique d'un échantillon est construite à partir d'un processus séquentiel combinant la transformation du signal, son filtrage et son codage. Des signaux synthétiques mimant les données chromatographiques ont également été utilisés.

4.2.1 . Préparation et analyse chromatographique

Les échantillons sont stockés à température ambiante dans des flacons en verre scellés. Ils sont utilisés dans l'état (sans lavage) avec une réduction possible si ils sont de taille supérieure à 1 cm de côté (produits finis). La durée de stockage entre la production et l'analyse a été inférieure à 24 mois pour tous les échantillons.

Les échantillons ont été préparés :

- par extraction solide-liquide suivi d'une reconcentration en vu d'une analyse en GC-MS (composés volatils et semi-volatils)
- soit par simple thermodesorption pour une analyse en HS-GC-MS (composés volatils concentrés)

4.2.1.1 . Extraction, reconcentration et analyse en GC-MS

Tous les échantillons sont préparés en duplicita, les produits finis sont découpés en morceaux de 1 cm de côté tandis que les granulés sont laissés tels quels. Ils sont pesés précisément pour atteindre environ 1 g dans un flacon en verre de 20 mL. Une extraction est réalisée avec 20 mL d'un mélange de dichlorométhane et de méthanol (v/v 3 :1) pendant 48 h à température ambiante sous agitation orbitalaire. La solution est alors transvasée dans un autre flacon en verre de 20mL avec une étape de filtration sur filtre toupie en téflon de 0.2 µm si la solution est trouble. L'échantillon est ensuite évaporé à sec avec un évaporateur centrifuge sous vide en trois étapes détaillées dans la table 4.4. Les échantillons sont enfin reconstitués avec 1,5 mL d'une solution d'étalons interne détaillée dans la table 4.5.

TABLEAU 4.4 – Paramètres de l'évaporateur

N° étape	Température	Vitesse de rotation	Pression	Durée
1	30 °C	20%	400 mbar	30 min
2	30 °C	20%	300 mbar	30 min
3	30 °C	20%	100 mbar	2 h

TABLEAU 4.5 – Solution d'étalons internes dans du chloroforme deutéré pour analyse en GC-MS

Molécule	CASRN	Rôle	Etat	Masse	RT*	RI*
Chloroforme-d	865-49-6	Solvant	Liquide	90 g	/	/
1,4-diacétylbenzène	1009-61-6	Étalon	Solide	10 mg	776 s	1472
Hexabromobenzène	87-82-1	Étalon	Solide	10 mg	1998 s	2558
Acétophénone-d8	19547-00-3	Étalon	Liquide	10 mg	538 s	1072

*Temps de rétention (RT) et indices de rétention (RI) dans les conditions de la méthode

Les analyses sont réalisées en GC-MS de type Varian 431-GC. La partie chromatographie en phase gaz est équipée d'une colonne apolaire Varian FactorFour VF-5MS 30 x 0,25 x 0,25. L'analyse est réalisée avec 1 µL de solution en mode split (1 :20) à une température d'injecteur de 275 °C avec de l'hélium en gaz vecteur à un débit de 1.1 mL/min. Le programme de température du four commence par un plateau à 50°C pendant 4 min, suivi d'un gradient de température de +15°C/min jusqu'à 320°C (18 min) et se terminant avec un plateau à 320°C de 12 min. La partie spectrométrie de masse est un simple quadrupôle configuré en mode scan de 40 Da à 650 Da avec un temps d'acquisition de 0,15 s, un délai de solvant de 2,5 min, les température de ligne de transfert et de source à 250 °C.

La séquence d'analyse type est détaillée dans la Table 4.6, où l'analyse "C7C40" désigne une solution d'alcanes allant de l'heptane au tétracontane préparée dans du chloroforme. Ils sont utilisés pour recalibrer les chromatogrammes en indices de rétentions, "Blanc méthode" est une solution qui a suivi le protocole de préparation, mais sans échantillons et "Solvant" correspond à l'injection de chloroforme uniquement.

TABLEAU 4.6 – Séquence type en GC-MS

N° injection	Nom injection
1	Solvant
2	Solvant
3	Solvant
4	C7C40
5	Solvant
6	Blanc méthode
7	Solvant
8	Échantillon 1 répétition 1
9	Échantillon 1 répétition 2
10	Solvant
11	Échantillon 2 répétition 1
12	Échantillon 2 répétition 2
...	...

4.2.1.2 . Thermodésorption et analyse en HS-GC-MS

Les échantillons sont pesés en duplicita dans des tubes de 20mL pour atteindre précisément environ 1 g (les produits finis hors granulés sont découpés en morceaux de 1cm de côté), 10 µL d'une solution d'étalons internes décrite dans la Table 4.7 sont ensuite ajoutés.

TABLEAU 4.7 – *Solution d'étalons internes dans du méthanol pour analyse en HS-GC-MS*

Composé	CASRN	Rôle	Etat	Masse ou volume	RT*	RI**
Méthanol	67-56-1	Solvant	Liquide	50 mL	/	/
Benzène-d6	1076-43-3	Étalon	Liquide	10 mg	586 s	959
Mésitylène-d12	69441-16-3	Étalon	Liquide	10 mg	1473 s	1253
Acétophénone-d8	19547-00-3	Étalon	Liquide	10 mg	2364 s	1646

*Temps de rétention (RT) dans les conditions de la méthode

**Indices de rétention (RI) dans les conditions standards polaires

Les analyses sont réalisées sur un appareil de type HS-GC-MS. L'échantillonneur headspace est un ITEX-STD-V3.0 configuré avec une incubation à 150 °C pendant 15 min et d'une seringue de prélèvement maintenue à 100 °C. La partie chromatographie en phase gaz est équipée d'une colonne polaire Agilent SOLGEL WAX 6 de dimensions 60 x 320 x 0,5, la température de désorption est configurée à 250 °C avec l'injection d'un volume de 750 µL en mode split 20 :1, le gaz vecteur utilisé est de l'hélium et le four est programmé pour avoir un plateau à 40 °C pendant 10 min, une rampe de 4 °C/min jusqu'à 200 °C maintenue pendant encore 10 min. Le détecteur est un simple quadrupôle configuré à 150°C avec une source à 230 °C et scannant les ions de 34 Da à 300 Da.

La séquence d'analyse type est détaillée dans la Table 4.8, où l'analyse "EI" désigne l'injection d'un tube contenant la solution d'étalons internes sans échantillons et "Air" l'injection d'un tube vide.

TABLEAU 4.8 – *Séquence type en HS-GC-MS*

N° injection	Nom injection
1	Air
2	Air
3	Air
4	EI
5	Air
6	Échantillon 1 répétition 1
7	Échantillon 1 répétition 2
8	Air
9	Échantillon 2 répétition 1
10	Échantillon 2 répétition 2
...	...

4.2.2 . Génération de chromatogrammes synthétiques

Des signaux synthétiques ont été synthétisés sur le modèle des données chromatographiques. Ces signaux sont composés de trois parties distinctes : ligne de base, bruit (grenaille et thermique) et des signaux d'intérêt assimilés à des gaussiennes. La ligne de base est constituée de deux plateaux de hauteurs différentes et d'une rampe pour les relier,

4.2.3 . Codage de signaux chromatographiques

Les données chromatogrammes bruts obtenus en analyse GC-MS ou HS-GC-MS sont disposées initialement sur une grille discrète irrégulière. La régularisation de la grille se fait en deux étapes. Dans un premier temps les données suivant l'axe masse sur charge subissent un groupement par classe (binning) unitaire sur plage d'ions couverte par le spectromètre basse résolution (40 Da à 650 Da en GC-MS, 34 Da à 300 Da en HS-GC-MS). Puis les données sont régularisées suivant l'axe des temps de rétention en utilisant le premier centile des pas de temps bruts comme intervalle pour une interpolation linéaire permettant ainsi d'éviter les effets de dépassement (overshoot ou undershoot).

4.2.4 . Encodage en séquences monotones

4.2.4.1 . Justification

Nous rappelons notre objectif d'évaluation et d'orientation de flux de matières par rapport à des états de références entraînant un besoin de méthodologie pour la comparaison d'échantillons ainsi que la construction d'une base de données viable sur le long terme.

La construction d'une telle ressource nécessite alors de réfléchir à une façon efficace de stocker des données variables et parfois massives. La stabilisation de données chromatographiques consiste à filtrer les différents biais rencontrés et indépendants de l'échantillon analysé : ligne de base (lié au saignement de la colonne à haute température), bruit (de grenaille et thermique), variation des temps de rétention (vieillissement de la colonne). Pour les deux premiers biais, nous avons fait le choix d'utiliser la transformée en ondelettes avec l'ondelette de Ricker à plusieurs échelles, permettant de filtrer efficacement les parties affines de la ligne de base ne laissant que des artefacts minimes aux zones de transition, le bruit est principalement extrait à l'échelle la plus basse permettant d'extraire ses propriétés, mais filtré aux échelles supérieures. Afin de réduire l'impact du vieillissement des colonnes sur les déplacements des signaux, nous avons besoin d'autoriser un glissement des données d'intérêts, pour cela nous avons décidé d'observer nos signaux transformés sous forme de segments monotones pouvant être décrits par trois propriétés : monotonie, signe de départ et s'il y a un changement de signe.

4.2.4.2 . Règles de codage

Les ions d'un chromatogramme sont traités à tour de rôle, la transformée en ondelettes continue avec l'ondelette de Ricker est appliquée aux échelles $2^i \quad i = 0..8$ donnant ainsi neuf signaux pour chaque ion. Ces signaux sont ensuite décomposés en segments monotones auxquels sont attribués six lettres selon les propriétés décrites dans la Table 4.9.

TABLEAU 4.9 – *Table d'encodage des segments monotones*

Lettre	Croissant	Début positif	Coupe l'axe X
A	1	0	1
B	1	0	0
C	1	1	0
X	0	1	0
Y	0	0	0
Z	0	1	1

Chaque échelle de chaque ion est ainsi encodée comme une séquence de lettres associées à la largeur et à la hauteur des segments monotones de la façon suivante L_w^h où L est une lettre autorisée par la table d'encodage, w et h sont respectivement la largeur et la hauteur du segment monotone considéré, le chromatogramme complet est donc réduit à un ensemble de séquences de lettres pour chaque ion et chaque échelle de la transformée.

Propriété 4.1 (Caractéristiques d'un segment monotone) *Chaque segment monotone se voit associer un caractère signant pour le triplet de propriétés monotonie, signe de départ, changement de signe. Un segment est ainsi dépouillé de sa hauteur et de sa largeur conservées à part. Deux segments signés par le même caractère seront ainsi équivalents peu importe leur amplitude en largeur ou hauteur.*

4.2.5 . Comparaison d'échantillons

La distance entre deux chromatogrammes (et abusivement entre deux échantillons) A et B est obtenue après alignement des codes de chacun des IC issus des chromatogrammes de A et de B. La distance entre le chromatogramme de A et le chromatogramme de B est obtenue sur la base de l'extensivité des critères retenus (qui n'est pas une métrique euclidienne).

4.2.5.1 . Alignement des codes

L'alignement des séquences de caractères est réalisée à partir de la fonction "nwalign" sur MATLAB en version R2020b (fonction introduite dans ToolBox "Bioinformatic" de MATLAB à partir de la version R2006a, sans mise à jour depuis) avec l'identité comme matrice de score pour favoriser l'alignement de caractères identiques.

L'alignement n'est pas réalisé manière globale mais IC par IC. Ainsi pour tout IC i de A et B, nous alignons A_i et B_i donnant un texte aligné $\phi_i^{A,B}$ indiquant les insertions P , suppressions M et substitutions S nécessaires pour passer de A_i à B_i et inversement ainsi que les invariance V . $\phi^{A,B}$ est défini comme la concaténation pour tout i des $\phi_i^{A,B}$.

4.2.5.2 . Critère de Levenshtein

La distance de Levenshtein définie par l'Équation 2.6 est généralisée aux chromatogrammes encodés par l'utilisation d'une somme entre les IC.

$$d_i^{A,B} = P_i + M_i + S_i$$

$$d^{A,B} = \sum_i d_i^{A,B} \quad (4.1)$$

4.2.5.3 . Critère de distance en entropie

Nous appelons dans ce travail "information mutuelle" les signaux des ions se ressemblant entre deux chromatogrammes A et B. Nous rappelons que pour tout IC i , A_i désigne le code de l'ion i dans le chromatogramme A, et $\phi_i^{A,B}$ est l'alignement des code de l'ion i entre A et B. Nous définissons l'information mutuelle pour un ion i entre A et B sur la base la séquence de caractère $\phi_i^{A,B}(V)$ où V contient les indices caractères inchangés entre A et B lors de l'alignement. $\phi_i^{A,B}(V)$ est nécessairement plus court que les codes A_i et B_i .

Nous définissons $P_\alpha(X)$ la fréquence d'occurrence du caractère α dans la chaîne de caractères X . Ainsi l'information contenue dans un des IC i des chromatogrammes A ou B est définie par :

$$H(A_i) = \sum_{\alpha} P_\alpha(A_i) \cdot \log_2(A_i) \quad (4.2)$$

Et l'information mutuelle entre A et B pour un IC i est définie par :

$$H(A_i \cap B_i) = \sum_{\alpha} P_\alpha(\phi_i^{A,B}(V)) \cdot \log_2(\phi_i^{A,B}(V)) \quad (4.3)$$

La distance en entropie est définie par l'entropie en excès par rapport à l'information mutuellement présente. Celle ci est définie IC par IC entre A et B par l'Equation 4.4 :

$$d_i^{A,B} = H(A_i) + H(B_i) - 2H(A_i \cap B_i) \quad (4.4)$$

Nous généraliserons cette distance aux chromatogrammes complets par sommation :

$$d^{A,B} = \sum_i d_i^{A,B} \quad (4.5)$$

5 - Résultats et discussion

Les résultats sont présentés dans cinq parties. Les deux premières parties présentent les résultats qualitatifs qui appuient les choix méthodologiques retenus dans cette thèse et qui sont aujourd’hui déployés au LNE (Laboratoire National de métrologie et d’Essais) pour des besoins industriels de gestion des risques sanitaires. Il s’agit de justifier la construction inductive de la méthode de traitement et de codage des signaux. Les interprétations classiques des signaux GC-MS se basant sur les corrélations entre ions pour caractériser des substances ne sont pas utilisées dans notre analyse. Il n’y a ainsi pas *d’*a priori** sur la nature et la présence des substances ainsi que sur la qualité de la concordance avec des substances connues (contenues dans des bases de données). A contrario, les signaux de GC-MS basse résolution seront considérés comme un signal temporel multiplexé incluant 611 ou 311 lignes de mesures, suivant que l’on soit en injection liquide ou en espace de tête. L’évolution temporelle dépend de la colonne et du programme de température utilisés. Les résultats présentés correspondent aux choix techniques du LNE utilisés en routine. Ils peuvent donc être considérés comme représentatifs des méthodes utilisées par des laboratoires de contrôle ou les acteurs du recyclage mécanique des matériaux.

La transformation des signaux de GC-MS par l’ondelette de Ricker conduit à de nouveaux signaux qui dépendent de l’échelle retenue ℓ . Il sera montré dans la Partie 5.2 qu’il est possible en choisissant de manière *ad hoc* les échelles d’isoler le bruit (signal chromatographique non significant) des signaux d’intérêt. Ces résultats réutilisent la propriété de stationnarité (au sens statistique) du bruit qui peut être aussi bien éliminé par le critère de rejet de la Partie 5.1 ou de manière plus robuste par une ondelette d’échelle ℓ optimale dans la Partie 5.2. Une empreinte globale des substances présentes dans l’extrait est proposée en choisissant une ondelette optimisée pour le programme d’élution retenu par le LNE. Le marquage des segments significatifs couplé à la monotonie de la réponse du détecteur garantit *a priori* l’existence d’une empreinte unique. Cette capacité est d’abord discutée en termes généraux pour séparer un PET vierge apte au contact alimentaire (REF G10-7) d’un PET recyclé d’origine mixte (REF G13-1). Cette analyse est toutefois suffisante pour conclure sur la capacité de la méthode à isoler un défaut sur une série de production ou à trier des gisements. Ces possibilités sont analysées sur l’ensemble des 165 échantillons réels dans les parties 5.3.2 et 5.4. L’utilisation d’un plus grand nombre d’échantillons permet de construire une base de comparaison dans l’espace des échantillons. Ce point de vue est complémentaire à l’analyse statistique des deux premières parties et intègre des biais autres que ceux associés au détecteur.

La dernière partie 5.5 réconcilie les deux points de vue notamment en présentant une approximation quantitative de la transformée inverse optimisée pour chaque ion. Parce que l’ap-proche complète n’utilise et ne détruit pas l’information chimique (corrélation entre les signaux temporels), les corrélations temporelles entre l’apparition des ions peuvent donc être utilisées pour identifier et quantifier les composés inconnus présents. Les possibilités d’identification des composés de manière peu ou pas supervisée sont particulièrement discutées et comparées avec d’autres méthodes d’analyse à plus haute résolution. La construction d’une base de données d’empreintes est esquissée tout en soulignant les limites des approches actuelles, dont le fait que les mesures soient issues d’un seul laboratoire.

5.1 . Modélisation du bruit associé aux signaux chromatographiques

L'approche méthodologique est justifiée sur des signaux chromatographiques typiques de GC-MS collectés par le LNE et appliqués à des extraits par solvant pour des matières plastiques. Ils sont représentatifs des mesures obtenues pour des composés semi-volatils issus de matériaux plastiques ou autres (ex. matériaux cellulosiques). L'approche est très générique et est appliquée de la même manière aux signaux obtenues en espace de tête pour des molécules volatiles extraites par thermodesorption. La démarche scientifique suivie est inférentielle. Elle repose sur un petit nombre d'hypothèses qui sont utilisées tant qu'une contradiction n'est pas identifiée. Les résultats méthodologiques sont ainsi découpés en deux parties, correspondant à un ou plusieurs choix de la méthode. La Partie 5.1.3 vise à construire un modèle statistique du détecteur qui permette d'identifier le rapport signal sur bruit en fonction du programme de température et de séparer ainsi les signaux signifiants (composés élusés) de ceux associés à la contamination de la colonne ou au vieillissement de la colonne. Ce critère n'est pas utilisé pour filtrer le signal, mais pour justifier les propriétés de filtrage (amplification) de la transformée en ondelettes de Ricker présentées dans la Partie 5.2.

5.1.1 . Temporalisation des signaux chromatographiques

Malgré le petit nombre d'échantillons réels étudiés ici (3), la robustesse de la méthode s'appuie ici sur un signal très riche dont la statistique peut être établie sur plus $m = 611$ ions fois $t = 13 \cdot 10^4$ temps d'élution ($\approx 8 \cdot 10^6$ points) malgré une acquisition en basse résolution. Ce type de traitement est rendu possible ici parce que la nature chimique des signaux est complètement ignorée et seule la nature temporelle du signal multiplexé est considérée. Une substance élue est ici un concept abstrait associé à l'apparition simultanée d'un ou plusieurs segments monotones indépendamment de leur nombre et de la concentration de la substance considérée. Le nombre de substances n'est à ce niveau pas identifiable mais le nombre de segments codant pour l'information de ces substances est dénombrable et peut être comparé à celui d'un autre échantillon.

Le signal chromatographique est multiplexé, c'est-à-dire que chaque ligne de masse est lue séquentiellement (commutée à tour de rôle) de la plus faible valeur de MZ vers la plus haute avec une période d'acquisition de 0.25 ms comme montré sur la Figure 5.1. En lisant le signal de manière temporelle pour chaque ion (lecture orthogonale à l'acquisition), chaque point de mesure se trouve séparé du précédent et du suivant par une période 0.15 s égale à la durée de la lecture l'ensemble des MZ. Chaque signal vérifie donc aisément la règle de Shannon et garantit l'indépendance des intégrations (comptages) réalisés pour le détecteur.

Du fait du multiplexage, la durée d'intégration ne représente toutefois qu'une fraction de la période d'acquisition. La temporalisation des signaux IC est essentielle à la méthode chromatographique et elle sera appliquée en considérant qu'un chromatogramme en GC-MS est équivalent à m signaux empilés comme si les m acquisitions n'avaient pas été multiplexées. Dans la Partie 5.2, la même propriété sera utilisée pour coder le signal transformé en rajoutant un caractère de fin (\$) qui marquera le passage d'une ligne de mesure à une autre. Il n'y a, en effet, pas de continuité temporelle entre chacun des m signaux des rapports MZ.

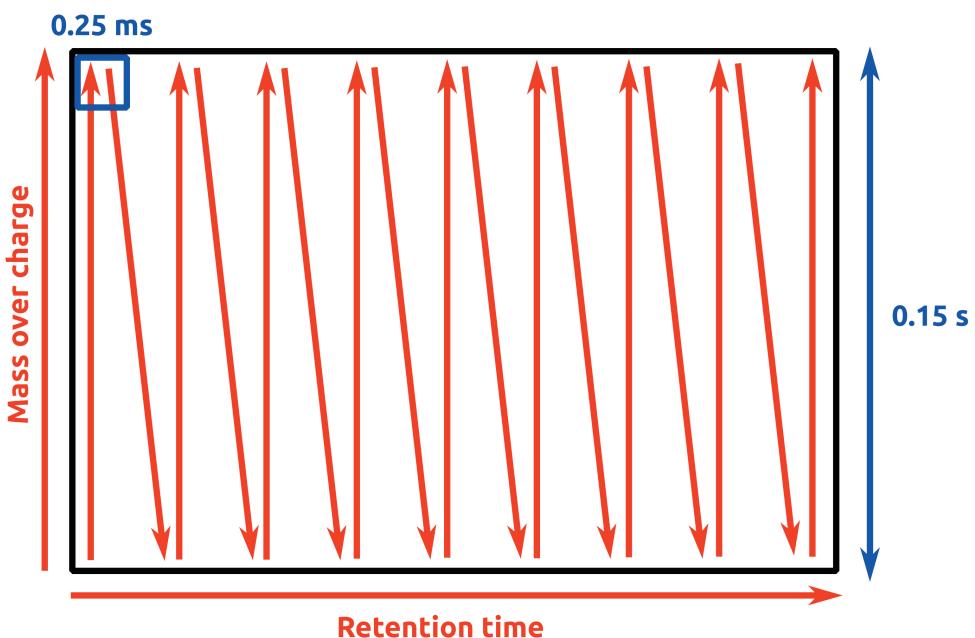


FIGURE 5.1 – Acquisition d'un chromatogramme par un appareil GC-MS avec un détecteur de type quadrupôle. Les rapports masse sur charge MZ sont acquis séquentiellement du plus léger au plus lourd avant de passer au temps de rétention suivant jusqu'à la complétion de l'analyse.

Points clefs

Nos signaux 2D typiques issus d'une analyse en GC-MS basse résolution contiennent 611 signaux temporels d'en moyenne 13000 points. Chacun des 611 signaux constitue un IC (Ion Chromatogram) dont la somme complète est le TIC (Total Ion Chromatogram). Les points sont acquis de manière séquentielle sur chaque IC avant de passer au temps de rétention suivant.

5.1.2 . Description d'un chromatogramme intégré typique

Le signal chromatographique GC-MS peut être visualisé soit comme un signal temporel (temps de rétention, RT) intégré sur la dimension du rapport masse sur charge (MZ), appelé "Total Ion Chromatogram" ou TIC, ou comme un signal brut 2D. La Figure 5.2 montre le signal TIC pour le programme de température standard du LNE pour les composés dits "semi-volatils". Ce programme comprend un plateau final à 320°C pendant 18 min qui garantit la désorption des molécules de masse moléculaires allant jusqu'à $551,5 \text{ g} \cdot \text{mol}^{-1}$ (par exemple l'hexabromobenzène). Les spécifications techniques de la colonne (modèle VF-5MS 30 m x 0,25 mm x 0,25 µm) mentionnent une stabilité jusqu'à 325 °C. L'échantillon contenait trois étalons (acétophénone-d8, diacetylbenzène et hexabromobenzène) qui correspondent respectivement aux temps d'élution 534 s, 772 s et 1193 s. Le même chromatogramme est présenté sur la Figure 5.3 en indice de rétention. La traduction des temps de rétention en indice de rétention est basée sur la calibration d'une série de 34 alcanes linéaires (C7-C40). Cette représentation qui ne sera pas utilisée dans le reste du document permet de comparer les résultats pour une colonne présentant une même phase mais des dimensions différentes et éventuellement un programme de température modifié. Les propriétés de la ligne de base ne sont toutefois pas affectées par la dilatation ou compression temporelle du signal.

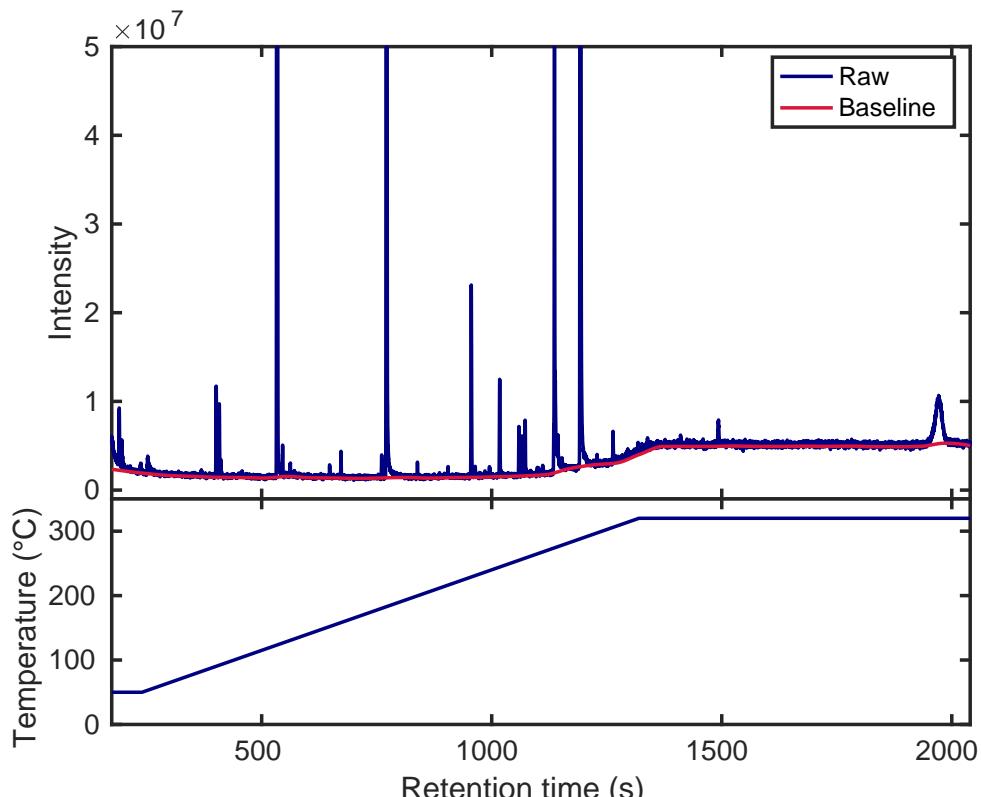


FIGURE 5.2 – Signal TIC typique d'un extrait de PET recyclé en GC-MS (13330 points) et programme de température du four. La ligne de base est définie comme la somme de la ligne de base de chaque ions, elle-même calculée à partir d'une médiane glissante centrée sur 501 points.

Les composés sont classiquement identifiés par des intensités supérieures à la ligne de base. On identifie 31 composés pendant le plateau de température à 50 °C et la rampe de température qui suit. À haute température, 3 composés sont décelables. Le signal de la ligne de base augmente significativement entre 1154 s et 1364 s (indices 2432 et 3158) et est associé à une élévation de température de 280 à 320 °C. Le maximum de la ligne de base est atteint dès que la température maximale est atteinte sans retard. Ce comportement souligne l'activation par la température des

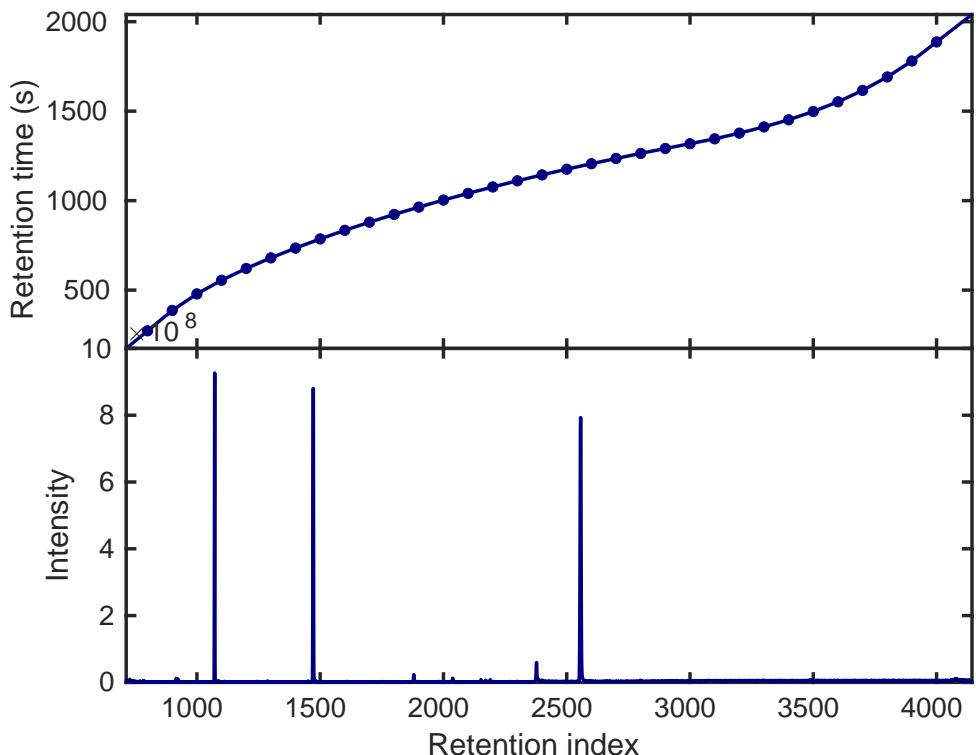


FIGURE 5.3 – Principe de la conservation du signal TIC de la Figure 5.2 en indice de rétention.

phénomènes contrôles la ligne de base. Le signal TIC bénéficie d'une bonne statistique (somme de 611 IC) et d'un codage en simple précision. Tous les nombres présentés sur les Figures 5.2 et 5.3 sont donc significatifs au sens de la mesure. Un seuil de coupure lié au matériel et au logiciel existe toutefois.

Le protocole utilisé a été standardisé pour permettre une identification générique des substances extraites, quelle que soit la nature du matériau (ex. polyesters, polyoléfines, polyamides, papiers, cartons), leur nature vierge ou recyclée. Il n'est pas donc pas adapté en fonction du type de substances attendues pour une situation donnée. Ces choix facilitent la comparaison des échantillons et le stockage des données sur la base d'un profil de contamination.

Points clefs

Le signal intégré le long de l'axe des IC aussi appelé TIC est constitué de deux sous-signaux majeurs :

- Des pics approximativement gaussiens avec une durée de vie limitée, et assimilés au passage d'un ou de plusieurs composés au détecteur après séparation par la partie chromatographie
- Une ligne de base bruitée corrélée avec le programme de température du four contenant la colonne chromatographique, cette ligne de base peut être estimée par une médiane glissante.

5.1.3 . Description du chromatogramme en deux dimensions correspondant

Les signaux IC 2D (611×13330) ne bénéficient pas de la loi des grands nombres (chaque point est accumulé sur 0.15 s), mais supporte l'information chimique : chaque substance est fragmentée en plusieurs ions caractéristiques. Les IC sont corrélés temporellement quand un composé donné apparaît. Les composés de l'extrait et les étalons internes sont associés à des signaux discrets à des temps d'élution donnés. Les composés issus de la dégradation de la colonne ou de sa contamination ont des réponses temporelles différentes et sont caractérisés par une plus grande persistance dans le temps. La ligne de base du signal TIC est calculée à partir de la somme des 611 lignes de base des 611 IC dont une majorité (361 sur 611) présente également une ligne de base significativement différente de zéro.

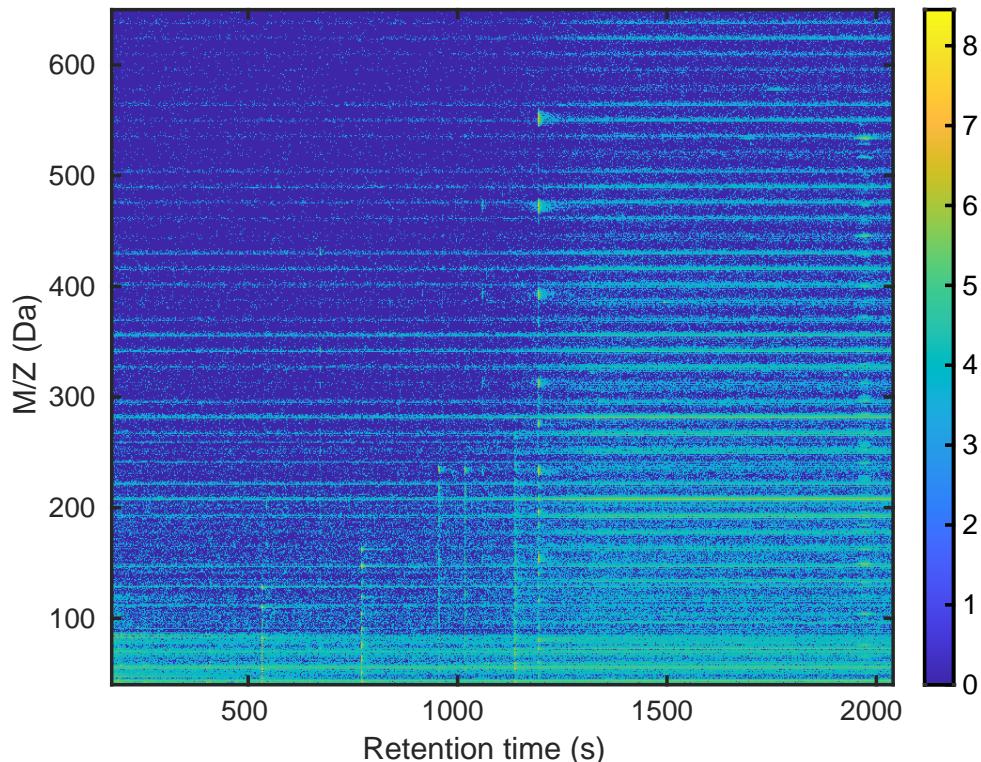


FIGURE 5.4 – Représentation en fausses couleurs du signal correspondant en 2D. Les temps de rétention sont en abscisse, les rapports masse sur charge en ordonnées et les intensités sont données en fausses couleurs selon une échelle logarithmique (base 10)

La Figure 5.4 montre l'évolution temporelle des 611 IC sous la forme d'une image en fausse couleur, où l'échelle de couleur (Parula allant du bleu au jaune) est associée à une échelle logarithmique de l'intensité mesurée. Les composés issus de l'échantillon sont associés à des stries discontinues verticales, qui correspondent aux impacts des ions de ces composés. Les lignes de base des IC apparaissent sous la forme de stries horizontales plus continues, c'est-à-dire sans retour à la couleur bleue. Il est important de noter que la résolution de l'image ne permet pas nécessairement de détecter les fluctuations temporelles de chaque IC. En supposant que la charge des ions est unitaire, le spot de masse moléculaire le plus élevé donne une indication de la masse moléculaire minimale des composés élusés. Cette analyse permet de justifier que la masse moléculaire des composés de l'extrait (y compris les étalons) augmente exponentiellement avec la durée d'élution : constante de temps de temps de 800 s pendant la rampe de chauffe. Les signaux IC 2D pourraient ainsi être décomposés en trois domaines :

- un domaine quasiment "vierge" de composés issus de l'échantillon et où seuls les signaux de la colonne peuvent exister
- un domaine où l'élution est activée par la rampe de température et où se mêlent les composés recherchés et ceux de la colonne
- un domaine complexe lié à l'exposition à haute température

En éliminant la zone de transition comprise entre 1154 s et 1364 s, une décomposition en quatre sous-domaines nommés par rapport aux points intercardinaux est préférée :

- SW (174 – 1154 s et 40 – 340 g · mol⁻¹) : zone activée par la température du four
- SE (1364 – 2040 s et 40 – 340 g · mol⁻¹) : zone des faibles masses à haute température
- NE (1364 – 2040 s et 340 – 650 g · mol⁻¹) : zones des fortes masses à haute température
- NW (174 – 1154 s et 40 – 340 g · mol⁻¹) : zone "vierge"

Points clefs

Les deux sous signaux observés sur le TIC (pics et ligne de base) existent de manière distincte sur chacun des IC. Les pics signant la présence d'une molécule apparaissent de façon synchronisée et forment la signature ou le spectre de la dite molécule. Les molécules les plus petites et polaires sont détectées en début de chromatogramme avec des spectres peu complexes concentrés sur les faibles masse. A l'inverse plus la température du four augmente, plus les molécules détectés sont grandes avec des spectres plus complexes et comportant des ions de masses élevées. Le comportement de la ligne de base est dépend de l'IC considéré, elle peut rester faible tout le long de l'analyse ou alors augmenter jusqu'à un plateau lorsque la température du four dépasse une certaine température.

5.1.3.1 . Modèles du détecteur et du signal mesuré $I_{t,m}$

Le détecteur (ici un simple quadrupôle) se comporte comme un compteur d'évènements discrets avec une période d'acquisition de 0.15 s . Pour chaque temps de rétention (RT), il compte séquentiellement les 611 fragments ayant un rapport masse sur charge (MZ) dans une gamme comprise entre $40\text{ g} \cdot \text{mol}^{-1}$ et $650\text{ g} \cdot \text{mol}^{-1}$ (temps d'acquisition Δt de 0.25 ms par valeur). Chaque évènement (i.e. ion arrivant au détecteur) active un multiplicateur d'électrons dont le gain en GC-MS est estimé à 10^5 d'après les documents du constructeur, cependant l'acquisition du signal au niveau du logiciel est très peu documentée et nous considérerons plutôt un gain effectif global dépendant de l'ion observé G_m . Ainsi la relation entre le nombre d'impacts et l'intensité du signal numérique peut s'écrire en considérant un gain effectif constant selon le temps de rétention et le rapport masse sur charge :

$$I_m^{tot}(t) = G_m \cdot c_m(t) \quad (5.1)$$

où $c_m(t)$ est le nombre total de coups ou d'ions de rapport masse sur charge $MZ = m$ à chaque instant t .

L'ensemble des évènements $c_m(t)$, quelle que soit la région considérée, peut être défini comme la superposition linéaire :

- des évènements liés à l'échantillon $s_m(t)$ (nul s'il n'y pas d'ion associé et approximé à une gaussienne sinon)
- d'évènements poissonniens associés à la dégradation ou contamination de la colonne dont le total est noté $b_m^{col}(t)$
- et éventuellement d'un bruit électronique $\epsilon_m^{T_{source}}(t)$

Le modèle statistique des évènements arrivant au détecteur s'écrit alors :

$$c_m(t) = s_m(t) + b_m^{col}(t) + \epsilon_m^{T_{source}}(t) \quad (5.2)$$

avec $b_m^{col}(t) \sim P(\lambda_m^{col}(t))$ où $P(\lambda_m^{col}(t))$ est la loi de Poisson dont le nombre moyen d'évènements par unité de temps pour chaque rapport masse sur charge m dépend $\lambda_m^c(t)$ est fonction de t .

Le gain effectif est estimé à partir du coefficient de variation $CV_m = \frac{\sigma_m}{\mu_m}$ de la série temporelle associée à la ligne de base de chaque ion $b_m^{col}(t)$ et sa comparaison avec le coefficient de variation théorique d'une loi de Poisson de paramètre λ : $CV_{Pois} = \frac{1}{\sqrt{\lambda}}$.

Ainsi le gain effectif G_m sur le signal colonne $I_m^b(t)$ peut s'écrire :

$$G_m = \frac{I_m^b(t)}{\lambda_m(t)} = I_m^b(t) \cdot \left(\frac{\mu_m}{\sigma_m} \right)^2 \quad (5.3)$$

où $I_m^b(t) = G_m \cdot b_m^{col}(t)$ correspond au signal colonne (évènements colonnes avec gain effectif).

Le gain effectif est fonction du rapport masse sur charge, cela peut être en partie expliqué par la variation du nombre d'électrons produits au niveau du multiplicateur en fonction de l'énergie de l'ion incident. Dans notre cas d'étude le gain effectif a été évalué pour chaque ion sur une plage de temps de rétention dépourvue de signaux issus de l'échantillon, celui-ci est compris entre 3000 à 10000 avec une moyenne approchant un facteur 5000 que nous conserverons pour l'ensemble du signal afin simplifier l'étude. Par la suite nous passerons d'un signal quelconque (total, échantillon, colonne) $I_m^{tot/s/b}(t)$ au nombre d'évènements correspondant en le divisant par le gain effectif G_m et la détection ou non d'un signal d'intérêt est équivalente à ce facteur près à la détection de l'évènement.

Points clefs

Le signal mesuré et enregistré dans les chromatogrammes est la mesure directe du nombre d'évènements (ions et bruit électronique) arrivant au détecteur multiplié par un gain (causé par le multiplicateur d'électrons servant à amplifier le phénomène physique en signal électronique) compris entre 3000 et 10000. L'ensemble des mesures peuvent être donc associées à des évènements de Poisson de paramètre λ , et en particulier le signal parasite issu de la colonne.

5.1.3.2 . Interprétation statistique des évènements liés à l'échantillon $s_m(t)$

L'échantillon étant constitué en partie de composés inconnus, son signal associé $s_m(t)$ n'est *a priori* pas connu. Il est reconstruit à partir d'une analyse statistique du signal colonne $b_m^{col}(t)$. Le signal échantillon interprété $\hat{s}_m(t)$ est défini à partir de la condition au sens probabiliste $c_m(t) \geq b_m^{col}(t)$.

Le signal d'un fragment d'une substance de l'échantillon est décelable si $s_m(t) > k_m(t)\lambda_m(t) \cdot \Delta t$, où $k_m(t)$ est un facteur de sécurité dépendant du temps de rétention et du rapport masse sur charge. L'inégalité de Bienaymé-Tchebychev assure que le nombre d'évènements issus la colonne est à plus de 90% compris entre $\lambda_{t,m} \cdot \Delta t \pm \sqrt{10\lambda_m(t) \cdot \Delta t}$. Du point de vue de la recherche de composés NIAS présents dans le mélange à des faibles concentrations, le biais introduit par la colonne augmente considérablement le risque de faux négatifs. En première approximation, un composé est détectable s'il vérifie les inégalités équivalentes :

$$\begin{aligned} c_m(t) &> \lambda_m(t) \cdot \Delta t + \sqrt{10\lambda_m(t) \cdot \Delta t} \\ s_m(t) &> \sqrt{10\lambda_m(t) \cdot \Delta t} \end{aligned} \quad (5.4)$$

Les seuils de détection entre deux localisations, (t_1, m_1) et (t_2, m_2) , contrôlés par des valeurs différents de $\lambda_m(t)$, varient comme le rapport $\frac{\lambda_{m_1}(t_1)}{\lambda_{m_2}(t_2)}$. A titre d'illustration, un signal faible inférieur à la valeur moyenne de la ligne de base ($\lambda_m(t)\Delta t$) serait théoriquement détectable que si la ligne de base est localement très faible. La probabilité que la contribution de la colonne soit plus faible que le signal recherché $s_m(t)$ est bornée par la première inégalité de Chernoff donnée dans l'Équation 5.5. A l'inverse un signal supérieur à la valeur moyenne de la ligne de base pourrait être masqué par une augmentation locale de la contribution de la colonne au signal total dont la probabilité est bornée par la deuxième inégalité de Chernoff donnée dans l'Équation 5.6.

$$Pr(s_m(t) \geq n_m^{col}(t)) \leq \frac{(e\lambda_m^{col}(t) \cdot \Delta t)^{s_m(t)} e^{-\lambda_m^{col}(t) \cdot \Delta t}}{s_m(t)^{s_m(t)}} \quad \text{pour } s_m(t) < \lambda_m^{col}(t) \cdot \Delta t \quad (5.5)$$

$$Pr(s_m(t) \leq n_m^{col}(t)) \leq \frac{(e\lambda_m^{col}(t) \cdot \Delta t)^{s_m(t)} e^{-\lambda_m^{col}(t) \cdot \Delta t}}{s_m(t)^{s_m(t)}} \quad \text{pour } s_m(t) > \lambda_m^{col}(t) \cdot \Delta t \quad (5.6)$$

Les Équations 5.5 et 5.6 démontrent que la capacité à détecter un ion unique est affectée par la présence d'ions (de même rapport m/z) parasites émis de manière relativement continue par la colonne. Ce brouillage du signal d'intérêt $s_m(t)$ (élution d'une substance de l'échantillon et détection d'un de ses fragments) dépend de l'ion considéré (ligne de mesure) et de la température du four au moment de son émission. La Figure 5.5 représente cinq isocontours (0.05, 0.25, 0.50, 0.75 et 0.95) des inégalités de Chernoff, l'isocontour à 0.05 permet entre autre de définir un domaine à l'est-sud-est où le signal a une probabilité d'au moins 0.95 d'être au-dessus de la ligne de base, un domaine au nord-nord-ouest où le signal a une probabilité de 0.95 d'être noyé dans la ligne de base et enfin un domaine au sud-ouest incertain où la ligne de base et le signal peuvent être localement plus ou moins intenses.

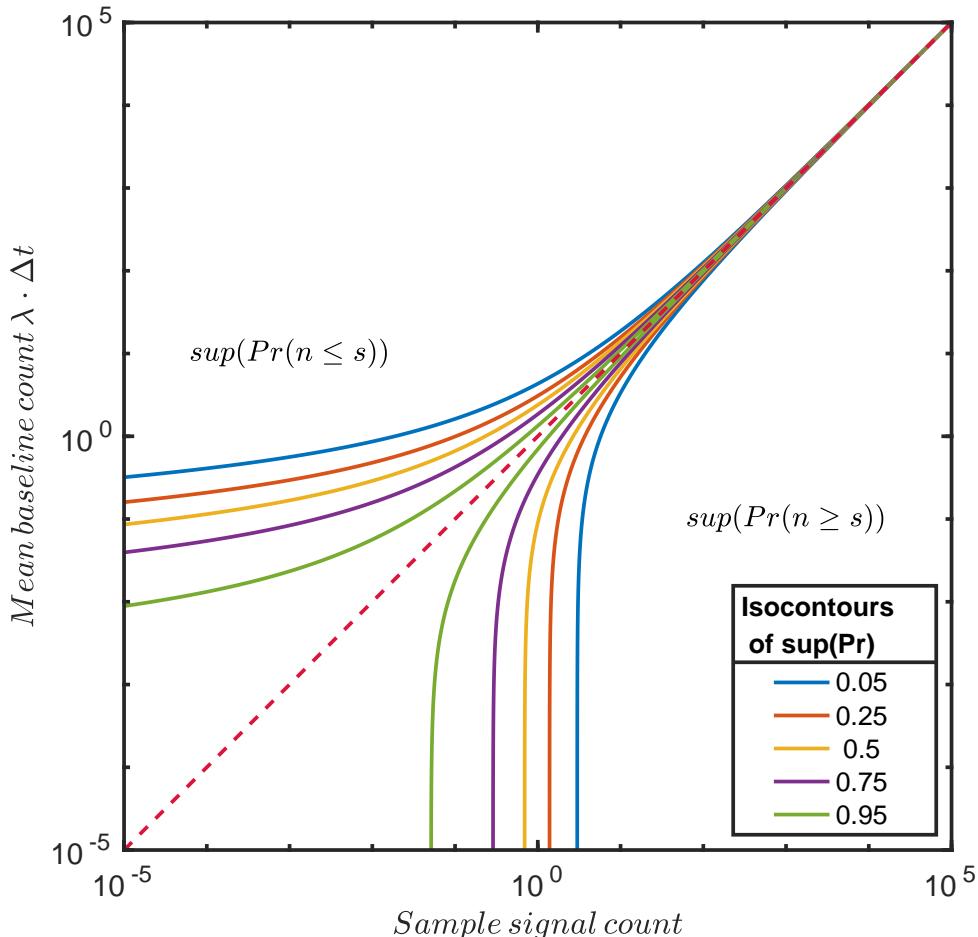


FIGURE 5.5 – Isocontours de l'inégalité de Chernoff entre un signal d'intérêt $s_{t,m}$ et une série d'événements poissonniens $n_m^{col}(t)$. La bissectrice coupe le domaine de probabilité en deux, lorsque $s_m(t) > \lambda_m^{col}(t) \cdot \Delta t$ la probabilité que le signal poissonnier soit plus important que le signal d'intérêt est bornée par le haut, à l'inverse lorsque $s_m(t) < \lambda_m^{col}(t) \cdot \Delta t$ la probabilité que le signal d'intérêt soit plus important que le signal poissonnier est bornée par le haut.

Par ce que la somme de lois de Poisson globalement indépendantes (pour des grandes populations) est aussi une loi de poisson dont le paramètre est la somme des $\lambda_m^{col}(t)$, nous en déduisons que la détectabilité d'une substance dépend de l'intensité de l'émission de composés par la colonne, mais surtout du nombre de ces composés. En spectrométrie de masse basse résolution, la capacité de détection est avant tout limitée par l'utilisation d'un grand nombre de lignes de mesures par la colonne. En d'autres termes, le problème de détection est similaire sur le signal cumulé TIC et obéit également aux Équations 5.5 et 5.6. Nous retiendra qu'il est globalement plus difficile puisque le cardinal des signaux échantillon $s_m(t)$ non nuls est statistiquement plus faible que le cardinal de signaux colonne $b_m^{col}(t)$.

Points clefs

Le signal lié à l'échantillon $s_m(t)$ même n'étant pas connu à l'avance, il est plus simple de traiter le signal résiduel causé par la méthode analytique et plus particulièrement par le saignement de la colonne $b_m^{col}(t)$. Nous avons fait l'hypothèse précédemment que le signal colonne est poissonnien de paramètre $\lambda_m^{col}(t)$. Sur cette base nous avons construit deux critères statistiques indépendants de l'échantillon pour discerner le signal causé par l'échantillon $s_m(t)$ du signal colonne $b_m^{col}(t)$. Le critère de Bienaym -Tch bychev (Equation 5.4) est applicable dans tous les cas et pour toutes les distributions, le crit re de Chernoff (Equation 5.5) est performant pour les queues de distribution uniquement.

5.1.3.3 . DéTECTABILITÉ D'UN ION ISSU D'UNE SUBSTANCE À FAIBLE CONCENTRATION

La capacité à identifier le signal d'un ion issu d'une substance à faible concentration $s_m(t)$ quand ce dernier est proche du bruit de Poisson du signal colonne $b_m(t)$ est illustrée sur les signaux d'ions typiques et discutée par rapport aux intervalles de confiance de la loi de Poisson associée au signal colonne $b_m(t)$. Les signaux d'ions typiques sont présentés sur la Figure 5.6 pour les trois classes de comportements observés :

- biais de colonne faible ($\lambda_m^{col}(t)\Delta t \rightarrow 0$) sur toute la durée du programme chromatographique (ions 114 et 390)
- biais de colonne fort ($\lambda_m^{col}(t)\Delta t > 0$) sur toute la durée du programme chromatographique (ions 83 et 429)
- biais de colonne variant de faible à fort lors de la montée en température (ions 319 et 577)

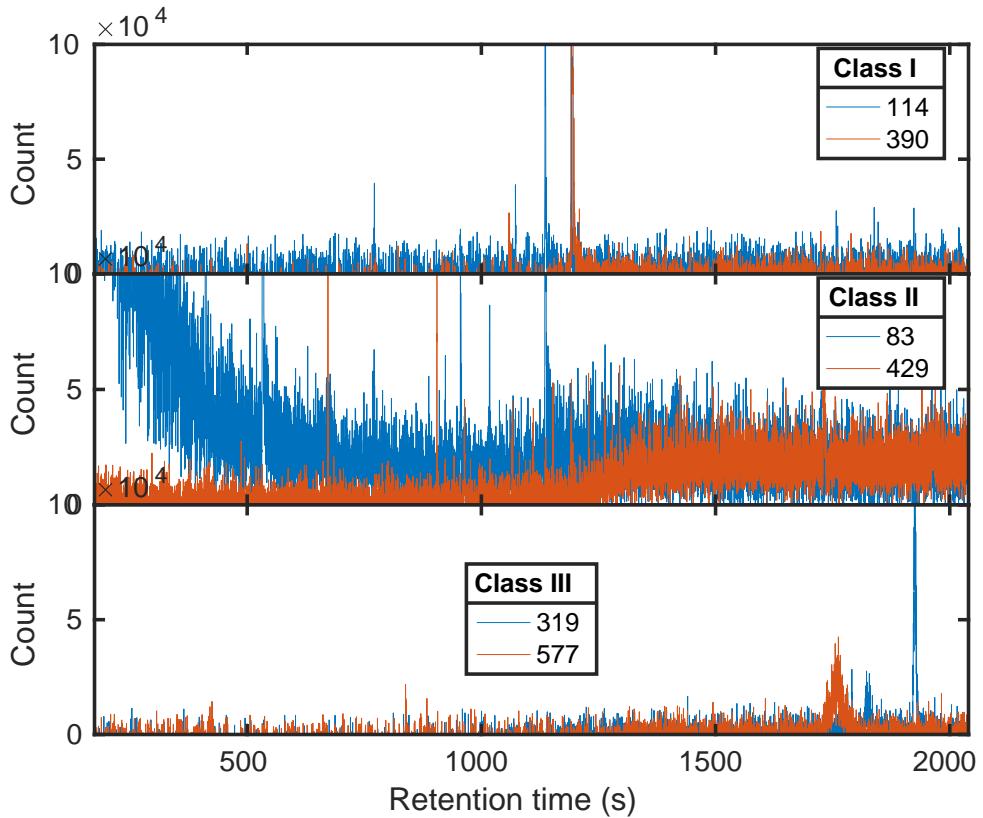


FIGURE 5.6 – Exemples de trois comportements de ligne de base

Le signal colonne $b_m^{col}(t)$ est estimé à l'aide d'une médiane mobile à 501 points appliquée au signal complet $c_m(t)$ permettant de filtrer les événements exceptionnels issus de l'échantillon $s_m(t)$, ce signal colonne estimé est noté $\widehat{b}_m^{col}(t)$. Ainsi le paramètre de la loi de Poisson estimé $\widehat{\lambda}_m^{col}(t)$ est décrit par l'Équation 5.7 :

$$\widehat{\lambda}_m(t) \cdot \Delta t \approx \frac{\widehat{b}_m^{col}(t)}{2} + \frac{\sqrt{225 \cdot \widehat{b}_m^{col}(t)^2 - 150 \cdot \widehat{b}_m^{col}(t) + 43}}{30} - \frac{1}{6} \quad (5.7)$$

Les Figures 5.7 et 5.8 montrent les chromatogrammes (TIC et 2D) correspondants aux données brutes de la Figure 5.4 après filtrage respectivement avec le critère de Bienaymé-Tchebychev (Équation 5.4) et le critère de Chernoff (Équation 5.6). Dans les deux cas, le contraste entre les signaux d'intérêt $s_m(t)$ et le signal de fond $b_m^{col}(t)$ est amplifié. Les ions qui apparaissent simultanément à la suite de l'élution d'une substance d'intérêt sont aisément discernables en accord avec le signal TIC déconstruit. Les biais résiduels sont associés à un bruit (des événements) globalement poissonnien de paramètre lambda allant de 0.05 à 110 pour la région basse température et de 0.05 à 192 pour la région à haute température. Le signal TIC en dehors des zones d'intérêt est lui-même poissonnien de paramètre lambda compris entre 154 et 660 pour la région basse température et entre 948 et 1115 pour la région haute température.

Nous remarquons que pour les ions présentés dans le Tableau 5.1, le nombre de composés détectables par les critères de Bienaymé-Tchebychev et Chernoff est surestimé, mais les deux approches restent cohérentes. Une grande partie des signaux $c_m(t)$ considérés plus grands que la ligne de base résiduelle sont très probablement des faux positifs et n'appartiennent pas aux signaux d'intérêt $s_m(t)$. On notera que l'information temporelle (forme et largeur des pics) n'est pas considérée dans cette approche tout comme la corrélation des ions au passage d'un composé d'intérêt. A titre de comparaison, un comptage manuel des composés sur le TIC seul conduirait à identifier 39 composés. La différence de décompte est associée au traitement statistique proposé. Le critère de Bienaymé-Tchebychev autorise un taux de faux positif jusqu'à 5%, il est normalement significativement plus faible. Le critère de Chernoff s'appuie sur un classement significatif – non-significatif meilleur qu'un classement aléatoire pour tous les pics retenus. Ce classement des signaux entre significatif (probablement faux positif) et non significatif reste objectif et est utilisable *a priori* avant ou après un traitement mathématique. Dans la suite du travail, la procédure statistique ne sera pas utilisée pour préparer les signaux de GC-MS, mais elle sera utilisée comme masque pour identifier les régions qui contiendraient des ions significatifs et qui devant être encodés avec la procédure proposée à l'aide de segments monotones. La transformée en ondelette permet en effet de s'affranchir de l'estimation de la ligne de base $\widehat{b_m^{col}}(t)$ à l'aide d'un filtre médian. Comme nous le verrons dans la Partie 5.2, la transformée de Ricker modifie le caractère poissonnien des événements de la colonne en imposant une moyenne nulle pour toute partie stationnaire tout en préservant sa variance initiale.

TABLEAU 5.1 – Application des critères de Bienaymé-Tchebychev et Chernoff pour la détection de zones d'intérêt à six ions issus des trois groupes de signal colonne

Type de signal colonne	Ion considéré	Nombre de segments d'après le critère opérateur †	Nombre de segments (largeur moyenne) d'après le critère BT*	Nombre de segments (largeur moyenne) d'après le critère C**
Biais faible	114	4	1077 (1.5)	1291 (1.7)
	390	2	513 (1.5)	659 (1.7)
Biais fort	83	5	184 (1.9)	18 (4.6)
	429	10	650 (1.6)	933 (1.8)
Biais variable	319	2	409 (1.6)	566 (1.7)
	577	2	113 (1.1)	290 (1.2)

† Pics sélectionnés sur le signal brut manuellement par un opérateur sur des critères de largeur et de rapport signal sur bruit

* Critère de Bienaym  -Tchebychev unilat  ral   95% de confiance

** Crit  re de Chernoff   50% de confiance

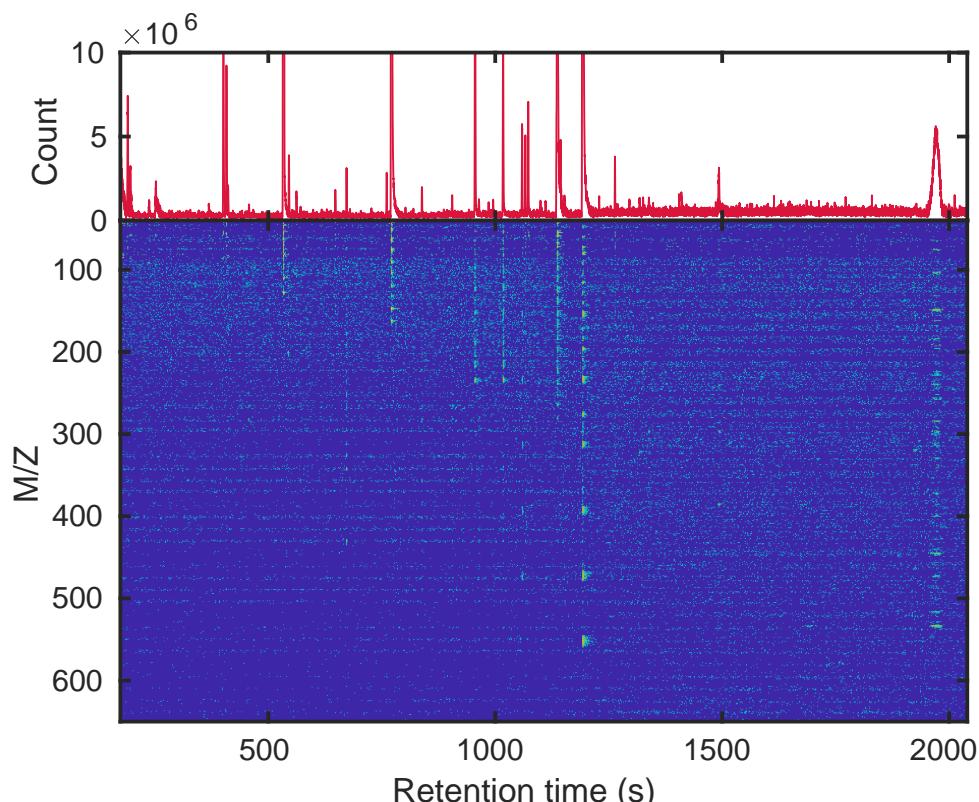


FIGURE 5.7 – Signal 2D correspondant filtr   avec le crit  re de Bienaym  -Tchebychev avec intervalle de confiance unilat  ral   5%

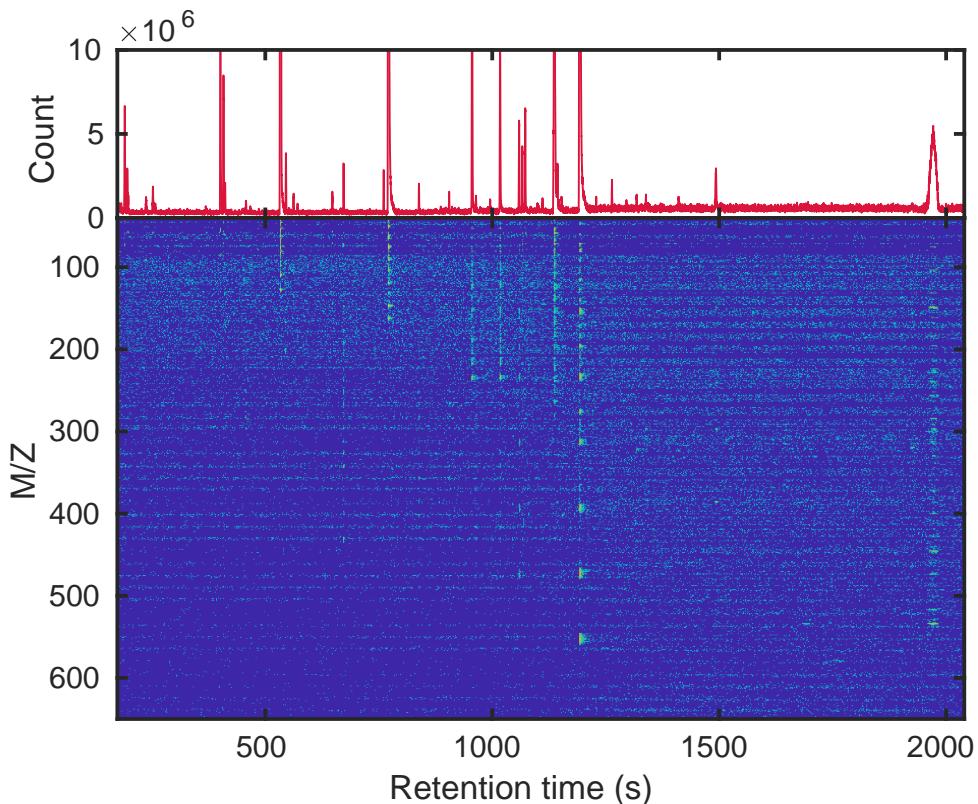


FIGURE 5.8 – Signal 2D correspondant filtré avec le critère de Chernoff avec intervalle de confiance unilatéral à 50%

Points clefs

Le signal colonne est estimé avec une médiane mobile large qui permet d'obtenir ensuite une approximation du paramètre de Poisson de chaque IC à chaque instant couplé à la fréquence d'acquisition de l'appareil connue à l'avance $\widehat{\lambda}_m(t) \cdot \Delta t$ à l'aide de l'Équation 5.7. Nous pouvons alors définir des filtres sur la base des critères de Bienaymé-Tchébichev et de Chernoff pour discriminer le signal signifiant $s_m(t)$ du signal colonne $b_m^{col}(t)$ et nettoyer les chromatogrammes sans supervision.

5.1.3.4 . Discussion sur la validité de l'hypothèse de Poisson pour le signal colonne $b_m^{col}(t)$

La séparation des signaux de la colonne $b_{t,m}^{col}$ et de ceux de l'échantillon $s_{t,m}$ pour un signal brut proche de la ligne de base repose sur une interprétation statistique et une distribution poissonniene des évènements de la colonne. Une condition nécessaire pour justifier une distribution de Poisson est de vérifier que la moyenne et la variance du nombre d'évènements associés à la colonne sont égales au paramètre λ . En tenant compte de la fréquence d'acquisition $f = 6Hz$ pour un IC, le critère de Shannon permet de construire une bonne statistique de la fréquence moyenne des évènements de la colonne jusqu'à une fréquence aussi petite que $\lambda = f/2 = 3Hz$. En deçà, il y a un risque significatif que le nombre d'évènements réels soient sous-estimés. La statistique doit être établie pour la zone non-isotherme et isotherme pour tenir compte d'une activation probable par la température.

L'hypothèse d'une distribution est testée sur la Figure 5.9 en traçant la variance en fonction de la valeur moyenne de l'estimation de la ligne de base $b_{t,m}^{col}$ pour les quatre régions étudiées (SW, SE, NE et NW). Des conclusions similaires sont obtenues à partir de l'injection d'un solvant pur qui ne contiendrait pas de substances. La relation linéaire entre la variance et l'estimation de l'espérance du biais confirme bien le caractère poissonnier de la réponse de la colonne au-delà de $\lambda = 10^{-4}$. La corrélation est plus approximative pour les fréquences inférieures notamment pour les ions les plus lourds à plus basse température (NW).

La dérive à une distribution de Poisson stricte peut être liée à l'estimation du signal colonne $b_m^{col}(t)$ par une médiane mobile supprimant les valeurs extrêmes et diminuant la valeur de la variance, certains signaux significants $s_m(t)$ peuvent ne pas être filtrés correctement.

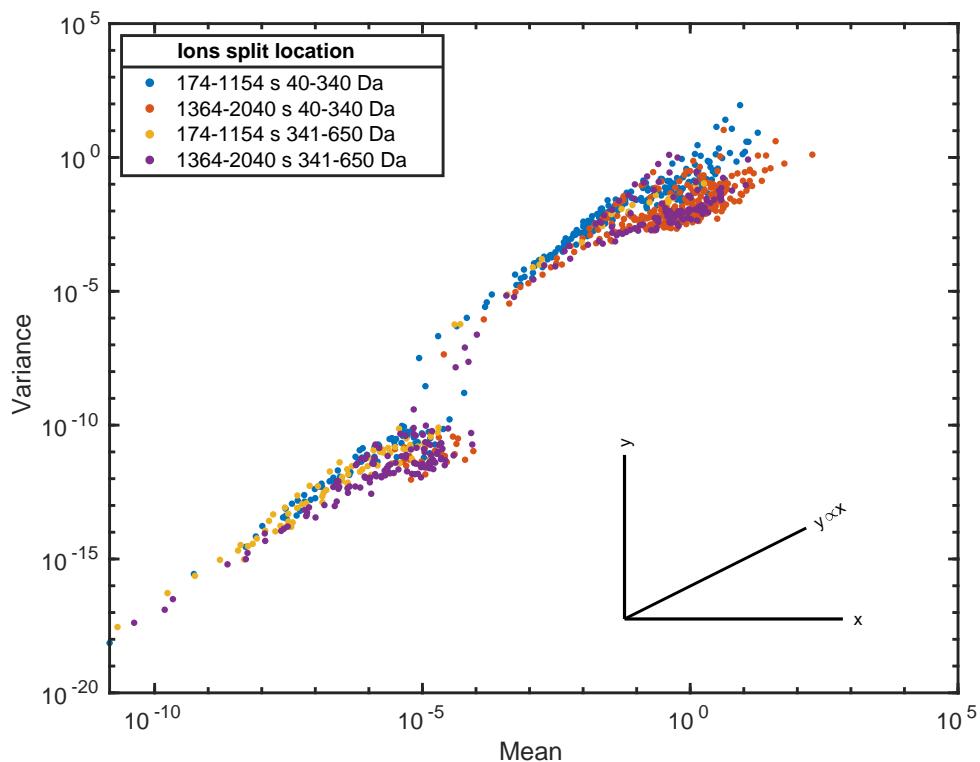


FIGURE 5.9 – Caractérisation des ions selon leur ligne de base et leur position

Points clefs

Les signaux de haute fréquence (paramètre de Poisson élevé) sont correctement évalués avec des moments de premier (moyenne) et second (variance) ordre égaux en accord avec l'hypothèse d'événements poissonniens. La cassure observée pour les signaux de basse fréquence est due à un échantillonnage insuffisant, conduisant à une sous estimation flagrante du nombre d'événements par unité de temps.

5.1.3.5 . Conclusions de l'analyse statistique du signal chromatographique brut

Le paramètre de la loi de Poisson est établi à partir d'une médiane mobile sur 501 points qui est sensée représenter la ligne de base $b_{t,m}^{col}$. Cette construction fait l'hypothèse implicite que le signal signifiant $s_{t,m}$ est minoritaire dans les événements collectés par le détecteur et n'affecte pas la position de la médiane et indirectement de la moyenne de la distribution. Cette approche est globalement satisfaisante sauf dans la zone d'activation du saignement de la colonne où le paramètre *lambda* évolue significativement. L'utilisation d'un modèle de Poisson pour le bruit confirme que le signal signifiant $s(t)$ ne peut pas être calculé comme la différence de $c_{t,m} - \langle b_{t\pm 250,m}^{col} \rangle$, où $\langle b_{t\pm 250,m}^{col} \rangle$ est le signal de la colonne estimé par une médiane sur 501 points (250 points à droite et à gauche d'un point central). D'une certaine manière, la différence précédente n'enlève, en effet, au mieux que 50% du bruit de Poisson de la colonne. L'originalité de l'approche proposée sur signal brut $c_{t,m}$ réside dans la prise de décision avec un critère de rejet du bruit au-delà de 50% pour minimiser le risque de faux positifs. Deux approximations des centiles élevés de la loi de Poisson ont été proposés sur la base des critères de Bienaymé-Tchebychev (BT) et de Chernoff (C). Les deux approches fournissent une borne supérieure à la probabilité que le bruit persiste au-delà d'un seuil retenu. Pour des événements rares qui ne sont pas bien échantillonnés, le critère BT est plus efficace et adapté même si le bruit n'est pas Poissonnier. Par exemple, pour $\lambda = 1$, BT prédit que moins de 25% du bruit persiste au-delà de $2 \cdot \lambda$ alors que C prédit 68%. Dans la suite de ce travail, une méthode de débruitage différente sera utilisée. Elle reprend l'hypothèse de stationnarité du bruit (événements indépendants) sans s'appuyer sur une hypothèse statistique quelconque. Cette opération est réalisée via une transformation continue (convolution) qui peut être réalisée très rapidement avec les processeurs arithmétiques présents dans les ordinateurs modernes. Le prix à payer est la transformation du signal qui perd son sens physique initial.

Points clefs

Dans la suite du manuscrit, nous ne conserverons que le critère de Bienaymé-Tchébychev (Équation 5.4) plus performant que le critère de Chernoff pour filtrer le signal colonne $b_{t,m}^{col}$. Ce filtre sera couplé à une transformation du signal afin d'améliorer ses performances.

5.2 . Transformée de Ricker et encodage des signaux chromatographiques

La transformation ajoute une dimension supplémentaire au signal initial 2D ($rt \times MZ$) avec l'échelle de la transformée qui permet de reconstruire plus ou moins indépendamment les détails (bruit) et les approximations à plusieurs échelles. En choisissant une échelle de référence, tout signal chromatographique devient encodable sous la forme de régions ou segments monotones le long de sa dimension temporelle. L'unicité de l'encodage ainsi la capacité à rapprocher ou distinguer deux échantillons seront discutés à la fin de cette partie.

5.2.1 . Transformée de Ricker

5.2.1.1 . Apport attendu de la transformation du signal

La transformée de Ricker offre une alternative non-déterministe au traitement statistique précédent. Il n'est pas nécessaire de déterminer quelle partie du signal contiendrait un excès d'évènement de la colonne ou pas. En effet, l'utilisation d'un noyau de convolution symétrique et non nul de type ondelette de Ricker (dérivée seconde d'une gaussienne) a trois effets attendus :

- le noyau annule ou atténue les signaux stationnaires (plateaux, rampes et signaux statistiquement stationnaires de type évènements de Poisson)
- le noyau se comporte comme un opérateur laplacien 1D et est capable d'extraire la dérivée seconde d'un signal transitoire symétrique
- la dérivée seconde du signal non stationnaire est lissée par une gaussienne de la largeur de l'ondelette de Ricker

Les deux dernières propriétés résultent des règles de différenciation des produits de convolutions à une variable. La seule propriété mathématique requise pour le signal continu d'intérêt est qu'il soit absolument intégrable (L1) et à dérivée faible. Ces propriétés mathématiques sont susceptibles de mieux poser le problème d'identification de composés de type NIAS à faible concentration sans utiliser les propriétés de corrélation des ICs, qui restent disponibles pour renforcer la puissance de l'analyse. Cette partie compare la capacité d'identification sur les mêmes signaux que ceux présentés dans la Partie 5.1.3.

Points clefs

La transformée de Ricker (convolution du signal avec la dérivée seconde d'une gaussienne dilatée à une échelle ℓ) a pour but de filtrer les signaux quasi stationnaires du point de vue de l'ondelette à l'échelle ℓ dont la rampe et le plateau liés au signal colonne $b_{t,m}^{col}$ à haute température. Le noyau gaussien contenu dans la convolution permet de lisser le signal et de réduire les variations du bruit résiduel, du point de vue spectral, cela revient à appliquer un filtre passe-bas sur le signal. Enfin le noyau contenant une dérivée seconde, la transformée permet d'extraire la courbure du signal.

5.2.1.2 . Signaux typiques transformés

Les signaux total $c_m(t)$, d'intérêt $s_m(t)$, issus de la colonne $b_m^{col}(t)$ et le bruit $\epsilon_m^{T_{source}}(t)$ transformés par l'ondelette de Ricker à l'échelle ℓ seront respectivement notés $C_{t,m,\ell}$, $S_{t,m,\ell}$, $B_{t,m,\ell}^{col}$, et $E_{t,m,\ell}^{T_{source}}$. Nous rappelons que la transformée en ondelettes est une application bilinéaire (Propriété 3.1), nous avons ainsi dans notre cas :

$$C_{t,m,\ell} = S_{t,m,\ell} + B_{t,m,\ell}^{col} + E_{t,m,\ell}^{T_{source}} \quad (5.8)$$

La Figure 5.10 présente les signaux des ions 83 (ligne de base variable) et 144 (ligne de base basse) et 319 (ligne de base haute) de la Figure 5.6 transformés par l'ondelette de Ricker aux échelles 1, 8 et 64. L'échelle correspond à la distance entre le maximum du lobe positif et le premier passage à zéro, le maximum et un minimum sont distants de 1.73 fois l'échelle, enfin l'ondelette vit jusqu'à une distance de 5 fois l'échelle depuis le maximum du lobe positif.

Les détails sont présentés dans la Partie 3.4.2 et visualisés sur la Figure 3.3. Il est important de noter que la transformée est réalisée sur signal discret sans utiliser de sur-échantillonnage. A l'échelle 1, le noyau de convolution contient 11 points non nuls. C'est le plus petit noyau de convolution applicable tout en préservant les propriétés générales de la transformée de Ricker. La forme du signal est affectée par l'échelle considérée mais avec les propriétés suivantes :

- A l'échelle 1, l'ondelette est la plus fine, le signal colonne s'exprime le plus mais sans sa ligne de base, les signaux de l'échantillon sont encore visibles mais grandement filtrés
- A l'échelle 8, le signal colonne est toujours présent, mais les signaux issus de l'échantillon profitent d'un gain en hauteur compris entre 1 et 2
- A l'échelle 64, les signaux sont encore plus lissés, mais le signal de l'échantillon perd en intensité par rapport au signal initial

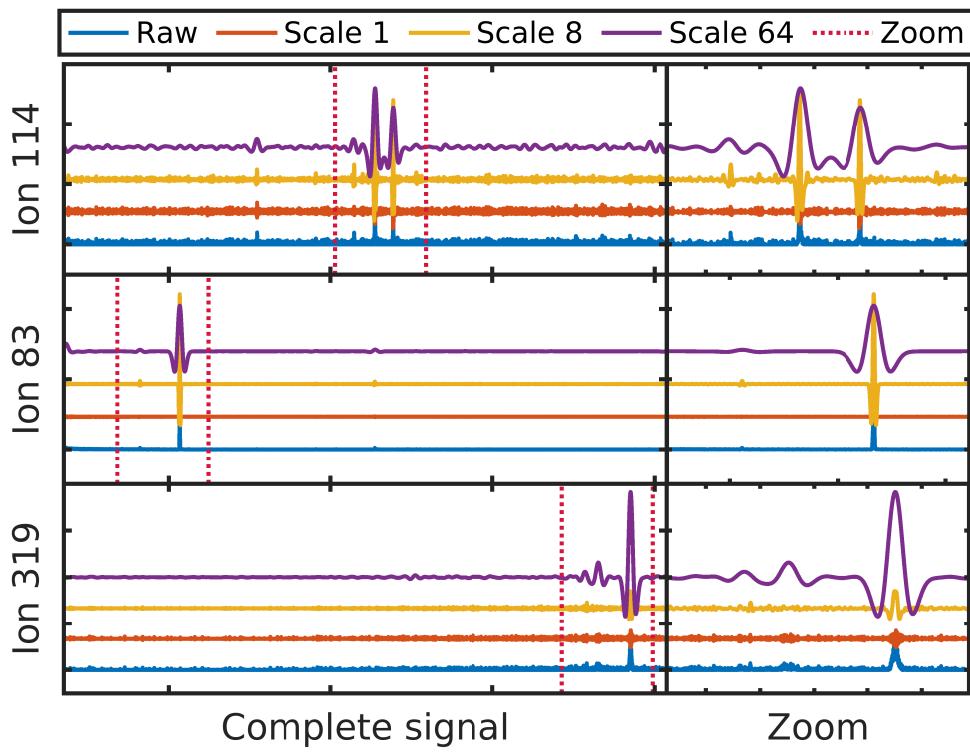


FIGURE 5.10 – Signaux brut et transformée en ondelette aux échelles 1,8 et 64 des ions 83, 114 et 319

Points clefs

La transformée de Ricker est une application bilinéaire, ses effets sur le signal complet peuvent être étudiés sur chaque sous-signal séparément. La transformation à plusieurs échelles crée un ensemble de données en trois dimensions $rt \times mz \times \ell$ avec une possible redondance entre les échelles ℓ .

5.2.1.3 . Explication de l'effet de loupe de la transformée sur le signal

Nous rappelons que l'utilisation de la transformée en ondelette ne modifie pas la variance des signaux infinis issus de variables aléatoires indépendantes. De ce fait, il ne s'agit pas d'une technique de filtrage, le bruit blanc résiduel n'est pas atténué. L'amplification des signaux d'intérêt ou l'effet de loupe résulte de la proximité des opérateurs mathématiques de convolution et de corrélation (les deux opérations sont identiques pour les noyaux pairs). Quand la largeur du lobe positif coïncide avec la largeur du signal d'intérêt $s_{t,m}$, il y a un renforcement du pic (voir la notion d'échelle optimale dans la Partie 3.4 et la Figure 3.4).

La position d'un pic isolé après transformée reste inchangée. Le signal d'une gaussienne d'écart-type σ transformé à l'échelle optimale ($\sigma_{opt} = 2.58 \cdot \sigma$) s'annule à $\pm 2.77 \cdot \sigma$ et est minimale à $\pm 4.79 \cdot \sigma$ autour de la position du pic. L'amplification maximale théorique (à l'échelle optimale) a pour valeur $1.588 \cdot \sqrt{\sigma}$. Contrairement aux méthodes statistiques de la Partie 5.1.3, la méthode de reconnaissance des ions après transformée de Ricker est capable de détecter des pics noyés dans le bruit de Poisson de la colonne à la condition que le signal du composé soit échantillonné sur une dizaine de points ou plus ($\sigma > 2$). Les amplifications obtenues pour les 13 pics visibles à l'échelle optimale théorique (en évaluant la largeur σ de chaque pic supposé gaussien) sont montrées sur la Figure 5.11 et comparées aux performances théoriques atteignables aux échelles $1 \cdot \sigma$, $2.58 \cdot \sigma$ et $4 \cdot \sigma$. Pour 10 des 13 pics, les résultats expérimentaux sont amplifiés comme prévus par la théorie avec une amplification \mathcal{A} comprise entre 2.2 et 3.5. Les trois autres pics sont amplifiés, mais avec des performances inférieures à la théorie du fait essentiellement d'une asymétrie prononcée.

Parce qu'un pic est détecté par un ensemble de points qui épouse la forme du lobe central de l'ondelette de Ricker, un pic devient décelable alors qu'il aurait été rejeté avec les critères statistiques de Bienaymée-Tchebichev et de Chernoff donnés respectivement dans les équations 5.4 et 5.6. En supposant une amplification \mathcal{A} moyenne de 3, ces mêmes équations permettent d'accepter des pics qui auraient été rejettés avec un risque de rejet de 55% au lieu de 95% pour Bienaymé-Tchébichev et un risque de 20% au lieu de 50% pour Chernoff. Il est rappelé qu'un critère de 50% correspond à un choix aléatoire des pics. La méthode proposée permet donc de réduire très fortement le risque de faux négatifs (signaux négligés à tort). Le problème des faux positifs est également amélioré pour les pics fins et isolés parce qu'ils sont associés à des valeurs $\sigma \leq 0.5$. Ils sont donc moins amplifiés et peuvent être éliminés en considérant des échelles optimales supérieures à $2.58 \cdot \sigma = 1.29$ (l'amplification \mathcal{A} théorique est au maximum de 1.11 alors qu'elle sera bien supérieure pour les pics plus larges). La résolution temporelle des pics devient alors un critère de filtrage.

Les variances des ions (découpés en régions basse température et haute température) de la Figure 5.6 avant et après transformée de Ricker aux échelles 1, 8 et 64 ont été compilées dans le Tableau 5.2. Si la variance du signal colonne $b_{t,m}^{col}$ est bien conservée à l'échelle 1 avec une différence maximale de 9% pour l'ion 577 à haute température, cette propriété n'est pas vérifiée sur les signaux typiques aux échelles supérieures (différences au-delà de $\pm 10\%$), car ils ne sont pas infinis et les signaux d'intérêt $s_{t,m}$ qu'ils contiennent ne sont pas indépendants. Cette particularité sera utilisée ultérieurement dans cette partie pour différentier les signaux d'intérêt $s_{t,m}$ des signaux colonne $b_{t,m}^{col}$ après transformée de Ricker.

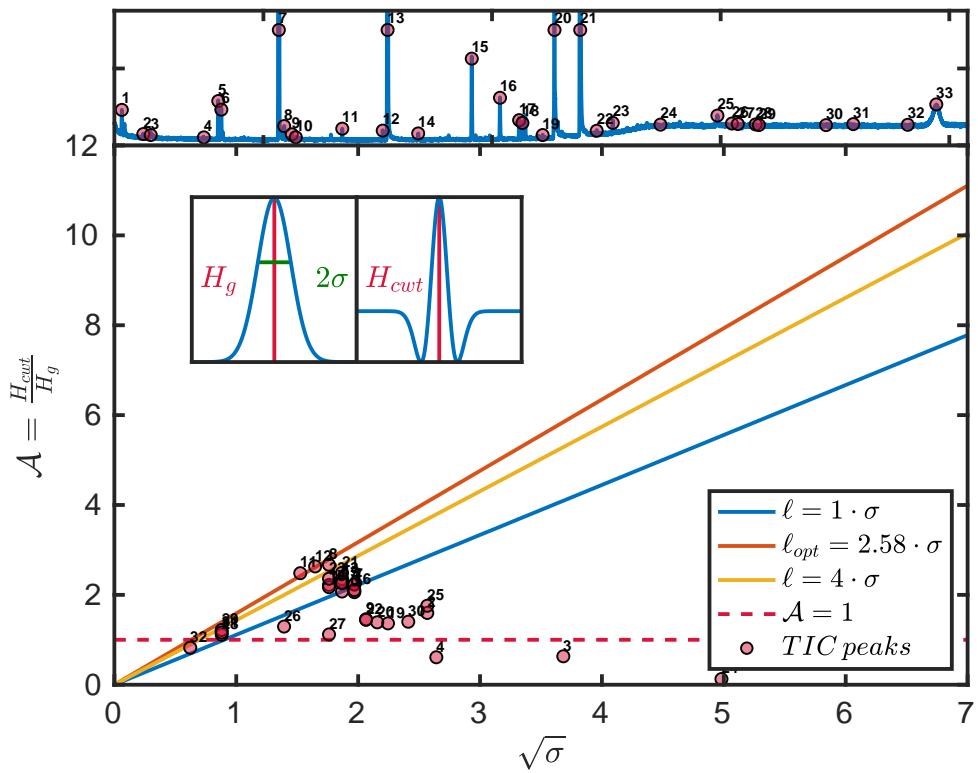


FIGURE 5.11 – Amplification maximale théorique en fonction de la largeur des signaux gaussiens et de l’optimalité de l’échelle de la transformée et comparaison avec signal réel

TABLEAU 5.2 – Variances du signal colonne $b_{t,m}^{col}$, à basses et hautes températures, brut et avec transformée en ondelette aux échelles 1, 8 et 64

Ion	Région*	Brut	Echelle $\ell = 1$	Echelle $\ell = 8$	Echelle $\ell = 64$
114	G	0.14	0.15	0.24	0.23
	D	0.28	0.30	0.46	0.44
390	G	0.02	0.02	0.03	0.04
	D	0.19	0.20	0.32	0.30
83	G	4.39	4.22	6.98	11.18
	D	3.77	3.96	6.50	7.31
429	G	0.48	0.48	0.87	1.22
	D	2.65	2.68	5.09	2.82
319	G	0.01	0.01	0.02	0.01
	D	0.10	0.11	0.19	0.06
577	G	0.01	0.01	0.01	0.03
	D	0.03	0.03	0.07	0.03

* G : Signal à basse température ($<300^{\circ}\text{C}$) ; D : Signal à haute température ($>300^{\circ}\text{C}$)

Points clefs

La transformée de Ricker, lorsqu'elle est appliquée à la bonne échelle $\ell = 2.58 \cdot \sigma$ où σ est la largeur d'un pic gaussien, permet d'amplifier les signaux d'intérêt $s_{t,m}$ selon un gain proportionnel à $\sqrt{\sigma}$. En parallèle la transformée ne modifie peu voire pas la variance du signal colonne $b_{t,m}^{col}$. Ces deux effets combinés permettent à la transformée d'augmenter significativement le rapport signal sur bruit lorsqu'elle est appliquée à une échelle ℓ optimale.

5.2.1.4 . Analyse statistique du signal transformé

L'utilisation de la transformée de Ricker n'empêche pas l'utilisation de critères statistiques pour raffiner l'analyse des signaux. Nous n'utiliserons ici que le critère de Bienaymé-Tchebychev unilatéral à 95% car celui-ci est applicable à n'importe quelle variable aléatoire contrairement au critère de Chernoff qui possède des formes spécifiques selon la distribution considérée.

Nous avons vu dans la Partie 5.2 que le signal colonne transformé $B_{t,m,\ell}^{col}$ a une espérance nulle et une variance égale à la variance du signal colonne brut $b_{t,m}^{col}$. Cependant nous n'avons à disposition que le signal total $c_{t,m}$, mais pour rappel d'après la Figure 5.10 à l'échelle $\ell = 1$ le signal colonne $b_{t,m}^{col}$ est quasiment le seul à s'exprimer : $C_{t,m,\ell=1} = B_{t,m,\ell=1}^{col}$. Cette observation est confirmée par le Tableau 5.2, la variance du signal colonne $b_{t,m}^{col}$ est proche de la variance de la transformée du signal total $C_{t,m,\ell}$ à l'échelle 1. Ainsi nous utiliserons cette évaluation de la variance $\lambda_{t,m}$ pour définir le critère de Bienaymé-Tchebychev par la suite. Ces spécificités sont résumées dans la Propriété 5.1.

Propriété 5.1 (Spécificités de la transformée de Ricker) *La transformée en ondelette de Ricker a des comportements significativement différents selon la nature du signal considéré. Le signal colonne transformé $B(t,m,\ell)$ possède une espérance nulle et une variance quasiment égale à celle du signal colonne initial $b(t,m)$. Les signaux d'intérêt $s_{t,m,\ell}$ sont amplifiés ou écrasés en fonction de l'échelle ℓ utilisée avec une amplification maximale obtenue pour $\ell = 2.58 \cdot \sigma$ où σ est la largeur du signal d'intérêt considéré. A l'échelle $\ell = 1$, le signal total transformé $C_{t,m,\ell=1}$ est quasiment égal au signal colonne transformé seul $B_{t,m,\ell=1}$ permettant ainsi d'extraire la variance du signal colonne et de l'utiliser pour différentier les signaux d'intérêt $S_{t,m,\ell}$ du reste avec un critère de Bienaymé-Tchébychev non supervisé.*

De ce fait le critère de Bienaymé-Tchebychev assure que le signal colonne transformé $B_{t,m,\ell}^{col}$ est à 90% compris entre $\pm\sqrt{10\lambda_{t,m} \cdot \Delta t}$. Le critère de Bienaymé-Tchebychev devient alors le même que nous considérons le signal total transformé $C_{t,m,\ell}$ ou le signal d'intérêt transformé $S_{t,m,\ell}$:

$$\begin{aligned} C_{t,m,\ell} &> \sqrt{10\lambda_{t,m} \cdot \Delta t} \\ S_{t,m,\ell} &> \sqrt{10\lambda_{t,m} \cdot \Delta t} \end{aligned} \tag{5.9}$$

L'application du critère de Bienaymé-Tchebychev sur les ions 114, 83 et 319 de la Figure 5.10 est illustrée dans la Figure 5.12. L'enveloppe proposée englobe le signal colonne $B_{t,m,\ell}^{col}$ transformé, les signaux transformés dépassant cette dernière sont considérés comme appartenant aux signaux d'intérêt transformés $S_{t,m,\ell}$.

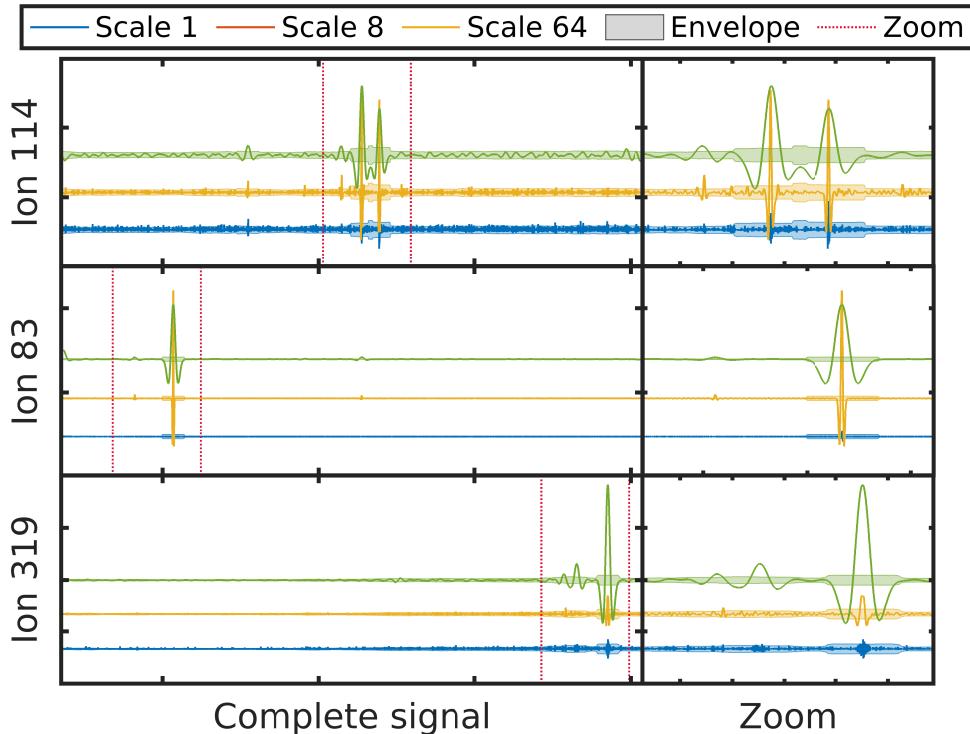


FIGURE 5.12 – Transformée en ondelette aux échelles 1, 8 et 64 des ions 83, 114 et 319 avec enveloppe du critère de Bienaym -Tchebychev

La comparaison de l’application du crit re de Bienaym -Tchebychev sur la base de la Figure 5.12 avant et apr s transform e de Ricker est compil e dans le Tableau 5.3. Nous rappelons qu’avant la transform e, de nombreux faux positifs de faible largeur (entre 1   2 points) taient dtect s. Apr s transform e, le nombre de signaux d’int r t $S_{t,m,\ell}$ dtect s est largement r duit (de plusieurs centaines  la dizaine de candidats). A l’chelle $\ell = 1$, la majorit  des signaux dtect s n’ont qu’une largeur d’un point en accord avec la propri t  de cette chelle  n’extraire quasiment que du bruit. A l’chelle $\ell = 8$ (optimale globalement pour nos signaux), le nombre de signaux d’int r t $s_{t,m}$ retenus est compris entre 36 et 48 avec une largeur moyenne de 10 points concordant ainsi avec le traitement manuel du signal (39 compos s). Enfin  l’chelle $\ell = 64$ le nombre de r gions d’int r t est grandement r duit, cela s’explique d’une part par le d passement de l’chelle optimale r duisant l’augmentation d’intensit  due  la transform e ainsi que par la fusion de pics proches (apr s convolution avec l’ondelette de Ricker).

TABLEAU 5.3 – *Comparaison de l’application du critère de Bienaym -Tchebychev unilat ral   95% de confiance avant et apr s transform e de Ricker pour la d tection de zones d’int r t   six ions issus des trois groupes de signal colonne*

Signal colonne	Ion	Nombre de segments (largeur moyenne)			
		Op�.†	Brut	Echelle $\ell = 1$	Echelle $\ell = 8$
Biais faible	114	4	1077 (1.5)	173 (1.1)	36 (7.2)
	390	2	513 (1.5)	157 (1.2)	45 (7)
Biais fort	83	5	184 (1.9)	37 (1.3)	43 (7)
	429	10	650 (1.6)	69 (1.1)	38 (7.9)
Biais variable	319	2	409 (1.6)	175 (1.2)	48 (7.3)
	577	2	113 (1.1)	192 (1.2)	46 (5.5)

† Pics s lectionn s sur le signal brut manuellement par un op rateur sur des crit res de largeur et de rapport signal sur bruit

Points clefs

La transform e de Ricker   l’ chelle minimale $\ell = 1$ permet d’isoler la contribution du signal colonne et d’en extraire la variance. A toute  chelle ℓ , la variance du signal colonne est inchang e  et sa moyenne est mise   z ro, seule la contribution des signaux d’int r t est modifi e .

Nous pouvons alors construire un crit re de Bienaym -Tch bychev dans l’espace transform  avec pour param tres une moyenne nulle et un estimateur fid le de la variance du signal colonne (Equation 5.9). Ce crit re permet de r duire grandement le nombre de faux positifs et faux n gatifs lorsque nous nous approchons de l’ chelle ℓ optimale.

5.2.2 . Encodage du signal transformé

5.2.2.1 . Principe et propriétés de l'encodage en segments monotones

La propriété de composition de fonctions monotones garantit une stabilité de l'encodage des régions croissantes ou décroissantes pour tout détecteur, qu'il soit utilisé dans son domaine de linéarité ou non. Il est donc possible d'encoder des signaux faibles même s'ils ne sont pas quantifiables. Comme cela a été montré dans la 5.2, la transformée en ondelette de Ricker est équivalente à la composition de deux transformations : un calcul de la dérivée seconde et un filtrage basse fréquence par une noyau de convolution gaussien. Le signe du segment code donc pour la concavité du signal vu à une échelle donnée. La concavité n'est pas détectable sur les pics chromatographiques de faible intensité, mais le devient après transformation si l'information de courbure est distribuée sur plusieurs points (si la durée de vie du signal est suffisante). L'utilisation d'une séquence pour identifier un signal d'intérêt (un ion) garantit une identification bien plus robuste que si elle était associée à la localisation d'un maximum. Les signaux chromatographiques indépendants (composés complètement isolés) et chevauchant (composés coélués) sont encodés différemment permettant ainsi de détecter la présence d'un composé même si ce dernier n'est pas identifiable.

Seules les propriétés de monotonie du signal transformé (positif et négatif) sont codées et utilisées pour identifier des différences entre signaux issus d'échantillons différents. La monotonie est une propriété mathématique capturée dans ce travail par le signe initial du signal, le signe de la variation, le changement éventuel de signe (ou le signe final), l'amplitude et la largeur (durée) de la variation sont aussi conservés. Les trois premières propriétés sont codées par des lettres et sont invariantes par rapport à l'intensité du signal et son éventuelle translation. A ce stade, il est important de remarquer que si la position du maximum (d'un pic isolé) n'est pas significativement déplacée à l'échelle ℓ de la transformation, la position du segment (ainsi que sa largeur et sa hauteur) sont très fortement affectés par le choix de cette dernière.

Le codage en lettres résume les variations (comme une table de variation) sous la forme d'une séquence compacte qui ne préserve ni les temps d'élution, ni les intensités. Il est possible en répétant les lettres (sur la largeur des segments monotones) de reconstruire un signal transformé qui préserve les positions des segments. Ce codage temporel ne permet pas d'obtenir une compression importante (autant de lettre que de points dans le chromatogramme), mais permet de comprendre comment le choix de l'échelle ℓ affecte le pouvoir de séparation de composés proches.

Le codage en lettres s'applique aussi bien aux grandes variations associées à l'élution des composés de l'échantillon qu'aux oscillations résiduelles associées à la transformée du signal de la colonne. Le codage de la monotonie des signaux bruts transformés $C_{t,m,\ell}$ conduit à la surreprésentation des lettres associées au signal colonne $C_{t,m,\ell}^{col}$ plutôt que du signal d'intérêt $S_{t,m,\ell}$. La Figure 5.12 démontre qu'une enveloppe de viabilité ou de confiance permet d'établir et d'identifier les zones d'intérêt séparées par des régions quasi constantes et sans intérêt.

Cette partie présente les principes et les propriétés du codage lettres avec ou sans filtrage. L'introduction d'une lettre supplémentaire " _ " (caractère vide ou blanc) permet la construction d'une séquence de lettres potentiellement unique pour un même échantillon malgré la présence d'un bruit résiduel aléatoire. L'unicité de la séquence est discutée en comparant la séquence de l'échantillon G5-1 avec sa répétition complète (protocole de préparation et d'analyse en double). L'information intensité n'est ici utilisée que pour éventuellement les segments monotones. Les pics chromatographiques de faible et forte intensité sont donc codés de la même manière (Propriété 4.1).

Points clefs

Le signal continu est transformé en séquences de caractères apparentés à un texte plus compacte. L'encodage de la monotonie du signal signal transformé permet de s'affranchir de l'intensité et de la largeur des signaux d'intérêt, mettant à égalité signaux proéminents et mineurs, mais aussi les segments résiduels du signal colonne transformé.

5.2.2.2 . Statistiques de l'encodage et critère de rejet du bruit

Le signal total transformé à l'échelle ℓ $C_{t,m,\ell}$ est encodé ligne par ligne (IC par IC), un caractère est attribué à chaque segment monotone sans distinction entre les signaux d'intérêt $S_{t,m,\ell}$ et les signaux issus de la colonne $B_{t,m,\ell}^{col}$. Les proportions des lettres cumulées sur les MZ sont indiquées dans la Table 5.4 pour le signal de la Figure 5.4 transformé aux échelles ℓ 1, 8 et 64. Les proportions des lettres ne sont pas uniformes pour les trois échelles ℓ . Les séquences sont majoritairement constituées de A et de Z à toutes les échelles ℓ , et en particulier à l'échelle $\ell = 1$ indiquant que les segments du signal colonne $B_{t,m,\ell}^{col}$ oscillent principalement autour du zéro. Nous remarquons aussi que les lettres de monotonies opposées (A et Z, B et X, C et Y) sont présentes en proportions égales. Enfin aux échelles supérieures, les proportions de lettres ne traversant pas le zéro (B, C et X, Y) augmentent en réponse à l'élargissement de l'ondelette et à la fusion de signaux d'intérêt $S_{t,m,\ell}$.

TABLEAU 5.4 – Proportions des segments monotones aux échelles 1, 8 et 64

ℓ	Nb. segments	A	B	C	X	Y	Z
1	5184872	0.45	0.01	0.04	0.01	0.04	0.45
8	577980	0.38	0.03	0.09	0.03	0.09	0.38
64	74273	0.41	0.03	0.06	0.03	0.06	0.41

Nous avons précédemment défini dans l'Équation 5.9 un critère de filtrage entre les signaux transformés d'intérêt $S_{t,m,\ell}$ et issus de la colonne $B_{t,m,\ell}^{col}$ basé sur l'inégalité de Bienaymé-Tchébychev et une évaluation de la variance du signal colonne transformé à l'échelle $\ell = 1$, $B_{t,m,\ell=1}^{col}$. Nous proposons de transposer ce critère en intensité par le critère suivant sur l'amplitude des segments traversants (A et Z) :

Proposition 5.1 (Critère de sélection des segments d'intérêt) *Un segment traversant (A ou Z) de hauteur $h_{A/Z}$ appartient au signal signifiant transformé $S_{t,m,\ell}$ si il vérifie l'inégalité :*

$$h_{A/Z} \geq 2\sqrt{10\lambda_{A/Z} \cdot \Delta t} \quad (5.10)$$

où $\lambda_{A/Z}$ est la variance maximale sur la largeur du segment du signal colonne évalué à l'échelle $\ell = 1$ $B_{t,m,\ell=1}^{col}$.

Nous rappelons que la transformée de Ricker change les distributions gaussiennes ou lorentzienne pouvant être définies par deux segments monotones en un signal à quatre segments monotones en commençant par un segment décroissant (X, Y ou Z), le filtrage doit alors conserver cette séquence minimale de segments consécutifs. Un pic isolé ou un massif de pics transformé commence nécessairement par un segment traversant descendant Z et se termine par un segment traversant montant A.

Le filtrage du signal encodé est un processus en plusieurs étapes. Les segments traversants (A et Z) sont filtrés par le critère de la Proposition 5.1 et remplacés par le tiret bas. Les segments non traversants (B, C, X et Y) sont ensuite filtrés si ils sont encadrés à gauche et à droite par des tirets bas. Les segments filtrés autour des segments restants sont restaurés de manière à avoir des longueurs de $2 + 2n$ avec n un entier strictement positif et des séquences commençant par un segment Z et se terminant par un segment A.

Le procédé de filtrage a été appliqué sur les segments de la Table 5.4 aux échelles $\ell = 1, 8$ et 64 , et les nouvelles proportions sont présentées dans la Table 5.5. Les séquences filtrées contiennent peu de segments d'intérêt, 4% à $\ell = 1$ en accord avec l'approximation $C_{t,m,\ell=1} \approx B_{t,m,\ell=1}^{col}$, la proportion de segments d'intérêt monte respectivement à 10% et 14% aux échelles $\ell = 8$ et $\ell = 64$. Cette augmentation peut s'expliquer par l'effet de loupe de la transformée en ondelettes aux échelles adaptées ainsi qu'à l'effet de coarse-graining de la transformée combinée à l'encodage en segments monotones.

TABLEAU 5.5 – *Proportions des segments monotones aux échelles 1, 8 et 64 après filtrage des segments non signifiant*

ℓ	Nb. seg- ments	A	B	C	X	Y	Z	_
1	5184872	0.02	0	0	0	0	0.02	0.96
8	577980	0.04	0	0.01	0	0.01	0.04	0.90
64	74273	0.05	0	0.02	0	0.02	0.05	0.86

Ce système d'encodage du signal chimique transformé a finalement plusieurs propriétés permettant de rendre similaires les signaux issus d'analyses successives d'un même échantillon et d'activer la comparaison non supervisée entre plusieurs signaux :

- Les lettres ne portent pas l'information en intensité du signal, des signaux d'intérêt issus d'une contamination ont alors le même poids que ceux d'un composé majoritaire

— Les segments de longueur variable sont chacun rassemblé sous un seul caractère

L'encodage permet ensuite l'utilisation d'alignement et de comparaisons de séquences.

Points clefs

Le critère de Bienaymé-Tchébychev (5.9) transposé aux segments monotones (5.1) permet de différencier les segments d'intérêt des segments résiduels à cause du choix d'encodage omettant les hauteurs. Le signal est désormais encoder avec six caractères signifiants (A, B, C et X, Y, Z) et d'un caractère signifiant pour le signal résiduel sans intérêt et considéré constant (tiret du bas _).

5.2.2.3 . Discussion de l'unicité de l'encodage entre répétitions d'un échantillon

La capacité à identifier deux échantillons comme identiques est testée sur l'échantillon de référence alimentaire (REF G10-7) et son injection répétée. La distance résiduelle entre les deux répétitions et son analyse statistique définissent la valeur d'un seuil minimal d'écart pour établir une différence significative. Afin d'obtenir des représentations graphiques qui permettent de repérer les zones différentes, l'analyse est d'abord réalisée avec les lettres répétées (séquences non compressées) pour la distance de Levenshtein (Équation 2.6) et l'entropie mutuelle (Équation 4.4). Les distances sont calculées pour chaque IC et sommées. Il s'agit donc d'une métrique L1 analogue des distances dites de Manhattan. Elle présente l'avantage de bien vérifier la loi faible des grands nombres tout comme le signal TIC et de pouvoir bénéficier d'un traitement statistique par échantillonnage (Monte-Carlo, Bootstrap , Jackknife) pour construire un intervalle d'incertitude. La notion de métrique s'appuie sur trois propriétés : symétrie, séparation et l'inégalité triangulaire. La distance de Levenshtein (Équation 2.6) par bloc constitue une métrique. La distance basée sur l'entropie mutuelle vérifie l'inégalité triangulaire à la condition que l'entropie calculée (Équation 4.4) soit positive ou nulle. Cette propriété devrait être acquise *a priori* pour des textes suffisamment longs, même aléatoires, mais des déviations pourraient apparaître sur des échantillons réduits. Aucune distance négative n'a été mesurée sur les échantillons réels étudiés.

L'unicité du codage est indispensable afin de pouvoir reconnaître le même échantillon quand celui-ci est injecté une seconde fois ou quand un matériau similaire est présent dans une série. En termes mathématiques stricts, on peut entendre le critère d'unicité soit comme une injection de l'ensemble des ions (indépendamment de leur intensité) présents dans l'échantillon vers les lettres codes, soit comme une mesure nulle de la distance entre deux échantillons identiques ou quasiment identiques. Le premier critère requiert une apparition des mêmes lettres exactement aux mêmes positions. Le second critère est plus générique et peut prendre en compte les erreurs de position et de lettres. C'est cette seconde approche qui est retenue, car le critère de rejet de bruit est statistique et laisse apparaître soit des lettres " _ " (blancs) ou des " fausses" lettres codantes. La notion de distance entre signatures peut être calculée soit pour un nombre de lettres imposé (les lettres sont répétées pour tenir compte la durée des segments), soit à partir des segments seuls indépendamment de leur longueur. Le dernier critère de distance est préféré dans ce travail parce qu'il inclut une notion de complexité du signal mesurable directement par le nombre de lettres. Cette notion de complexité présente certaines similarités avec le concept de complexité algorithmique ou de complexité de Solomonoff–Kolmogorov–Chaitin. Nous nous attendons intuitivement à avoir un signal plus complexe pour un matériau recyclé que pour un matériau vierge. Cette analyse n'est toutefois possible que si le critère de rejet du bruit a permis de simplifier suffisamment un chromatogramme qui ne contiendrait aucune substance d'intérêt.

L'étude précédente de l'encodage du signal et du filtrage sur un critère statistique dérivé de l'inégalité de Bienaymé-Tchebychev a montré que seule une portion réduite des segments monotones code pour le signal signifiant $S_{t,m,\ell}$ (10% à l'échelle optimale $\ell = 8$). Nous devons nous assurer que cette partie du signal constitue la signature de l'échantillon étudié tant par son analyse répétée. Pour cela nous étudierons l'encodage de deux répétitions de l'échantillon G5-1 (PET vierge). Dans cette étude l'encodage du chromatogramme est répété selon la longueur du segment monotone considéré afin d'obtenir une matrice rectangulaire de lettres (un segment monotone A de longueur 5 devient AAAAA).

La Figure 5.13 présente la comparaison des deux répétitions de l'encodage de l'échantillon de PET vierge G5-1. Le cadre inférieur gauche indique en bleu les lettres communes aux deux répétitions et en jaune les lettres différentes, nous pouvons observer des différences conséquentes à la périphérie des pics et dans la zone SW (basse température et haut). La présence de molécules communes est marquée par des traits bleus verticaux et une légère chute locale de la distance et de l'entropie par colonne.

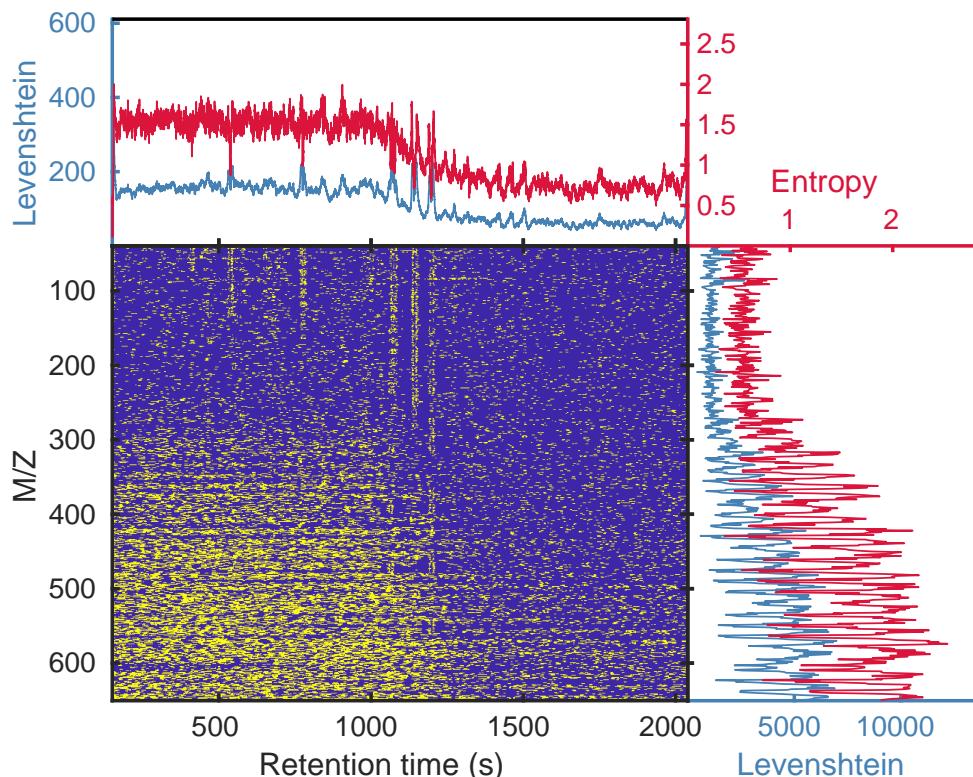


FIGURE 5.13 – Comparaison de l'encodage à l'échelle $\ell = 8$ des répétitions de l'échantillon de PET vierge G5-1

Points clefs

La répétition des lettres permet de comparer spontanément deux chromatogrammes sans aucun autre traitement. La comparaison de deux répétitions d'un même échantillon n'est pas exempte de défauts et souffre des variabilités liées à la préparation des solutions ainsi qu'à l'analyse en elle-même. Le bruit résiduel situé dans la zone basse température et hautes masses moléculaires, ainsi que le voisinage des pics contribuent pour la majorité de la distance.

5.2.2.4 . Discussion de la différence d'encodage entre deux échantillons

Nous allons comparer ici deux échantillons distincts, le même PET vierge de la partie précédente G5-1 et un PET recyclé non décontaminé G13-1 contenant bien plus d'information. La comparaison se fait toujours sur la base de l'encodage répété des chromatogrammes.

La Figure 5.14 présente la comparaison entre le PET vierge G5-1 et le PET recyclé non décontaminé G13-1. Le biais de la zone SW est toujours présent, les différences entre les deux échantillons sont marquées par un ensemble de tâches jaunes alignées verticalement signant pour la présence exclusive d'une molécule dans un des échantillons. La différence entre échantillons est aussi visible sur le cumul des ions, les molécules exclusives à l'un ou l'autre échantillon créent des pics où forêts de pics dont la hauteur est liée à la complexité du spectre de cette molécule. Ce signal temporel ressemble à s'y méprendre à un chromatogramme avec une ligne de base forte à gauche. *A contrario* le cumul temporel n'apporte pas d'information nette. Par rapport à la comparaison des répétitions de la partie précédente, les valeurs des deux critères (Levenshtein et entropie) restent élevées pour tous les ions.

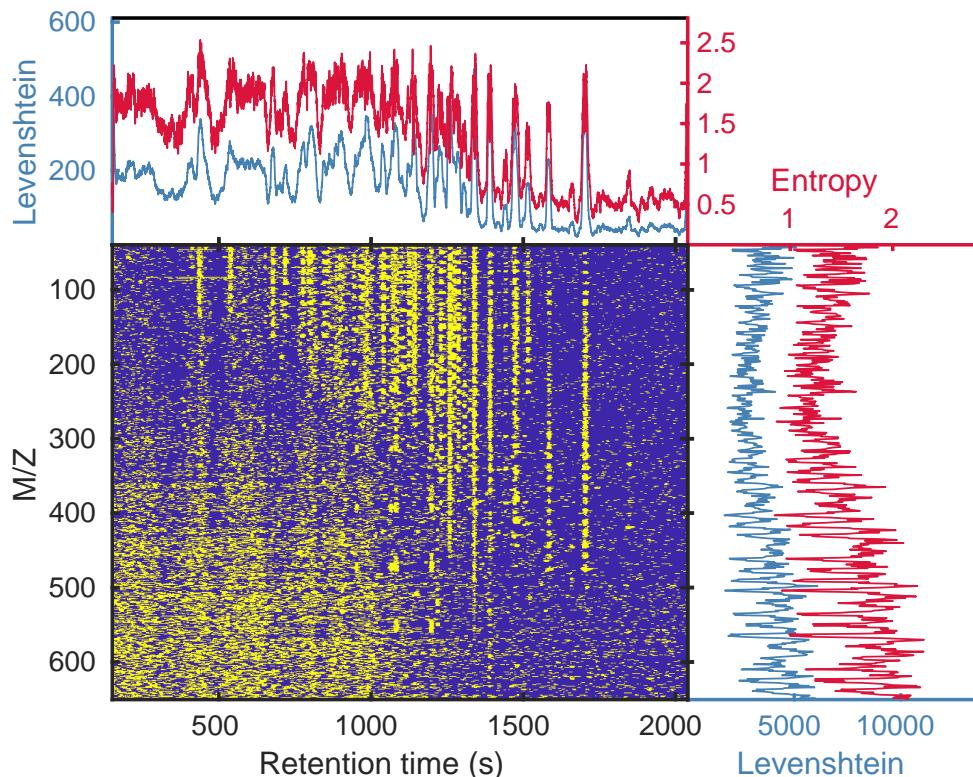


FIGURE 5.14 – Comparaison de l'encodage à l'échelle $\ell = 8$ de du PET vierge G5-1 avec le PET recyclé non décontaminé G13-1

Points clefs

Les différences de composition sont marquées par des stries jaunes verticales signant pour des composés exclusifs à l'un ou l'autre des échantillons. Le cumul des différences ligne par ligne assimilable à un TIC montre des pics à chaque temps de rétention où une différence apparaît. La zone vierge basse température et hautes masses moléculaires est toujours gangrénée par le bruit résiduel.

5.2.2.5 . Discussion sur les performances de la comparaison en caractères répétés et compressés

Les comparaisons des séquences avec lettre répétées permettent d'avoir des représentations proches des chromatogrammes initiaux, cependant celles ci sont sensibles au bruit de Poisson résiduel qui n'aurait pas été filtré par le critère statistique de Bienaym -Tch bychev et qui influence sur la largeur des segments. De plus la rép titions des caract res r duit dramatiquement la capacit  de compression de la m thodologie. Enfin le Tableau 5.6 donne les distances pour les crit res de Levenshtein et en entropie pour la m thode avec r p tition des caract res et pour la m thode compress e avec alignement. Le rapport des distances entre chantillons sur r p titions est un estimateur des capacit s  discerner correctement des chantillons r ellement diff rents. La comparaison des s quences compress es apporte de meilleures performances dans le cas de la distance de Levenshtein avec un rapport trois fois plus lev . Pour le crit re en entropie, le gain est minime, voire nul, pour l'exemple consid r . Cependant les autres apports de la compression comme la r duction de la sensibilit  au bruit de Poisson r siduel ainsi que la diminution de l'espace de stockage et de bande passante n cessaires pour conserver et transf rer des donn es.

TABLEAU 5.6 – *Comparaison de la discernabilit  des m thodes r p t es et compress es pour les crit res de Levenshtein et en entropie sur les r p titions de PET vierge (D_{rep}) et sur l'chantillon de rPET non d contamin  (D_{diff})*

M�thode	Crit�re	D_{rep}	D_{diff}	$\frac{D_{diff}}{D_{rep}}$
R�p�tition	LVS	$1.56 \cdot 10^6$	$2.10 \cdot 10^6$	1.35
	ENT	1.20	1.54	1.28
Compression	LVS	$6.89 \cdot 10^4$	$3.33 \cdot 10^5$	4.83
	ENT	0.49	0.66	1.35

La comparaison de chromatogrammes avec les caract res r p t s permet de simplifier son exploitation graphique par un op rateur. Mais cette approche n'est pas compatible avec un traitement non supervis  d'un large jeu de donn es, nous sommes oblig s de sommer l'ensemble des distances. Dans ce cas l'unique int r t de la r p tition s'efface au profit d'un scalaire caract risant l' cart d'informations entre deux chromatogrammes et nous avons pu voir que la capacit   discerner deux chantillons est meilleure pour les s quences compress es mais align es.

Points clefs

La comparaison de chromatogrammes avec r p tition permet de distinguer les points, temps de r tention ou ions signant pour la diff rence au cas par cas. Cette approche est int ressante pour un jeu restreint d'chantillons, car elle ne n cessite pas d'expertise ni d'exploitation pouss es  l'op rateur. Cependant la comparaison des s quences compress es et align es permet une meilleure discernabilit  (rapport entre la distance entre chantillons sur l'incertitude li e  la manipulation et  l'analyse) pour le crit re de Levenshtein et dans une moindre mesure pour l'entropie.

5.2.2.6 . Résumé de l'encodage et de la comparaison

Nous retiendrons essentiellement que la méthodologie proposée pour identifier des substances inconnues peut, contrairement à beaucoup de méthodes chimiométriques, opérer avec seulement deux chromatogrammes sans recourir à la moindre calibration. Nous supposerons seulement que les deux échantillons ont été obtenus avec un programme de température et une colonne comparables. La colonne peut avoir vieilli et donner lieu à des profils de saignement différents. L'hypothèse fondatrice est que le signal de la colonne (hors composés d'intérêt) est poissonien et peut être éliminé soit à partir d'un critère statistique de rejet sur le signal réel (Figures 5.7 et 5.8), soit d'une estimation du bruit sur le signal transformé à l'échelle $\ell = 1$ pour construire une enveloppe statistique à toute échelle ℓ (Figure 5.12).

Pour nos signaux GC-MS, l'échelle $\ell = 8$ permet d'amplifier naturellement les pics associés aux principaux composés volatiles et semi-volatiles issus des matériaux plastiques. Il s'agit de substances variées (monomères, additifs, résidus, contaminants) qui peuvent être coélues, mais qui sont *a priori* distribuées sur la plage d'élution optimisée par le programme de température du four. Le critère de rejet est appliqué sur les IC (ion par ion) et non sur le TIC complet (substance par substance). Si un des ions de la substance d'intérêt est différent du bruit généré par la colonne, il sera conservé. Tous les ions d'intérêt simultanés forment une signature de la substance. Ce choix est au cœur de l'approche proposée et est indépendant de notre capacité à quantifier la substance ou même à l'identifier. C'est la signature, mais tronquée et partielle qui est privilégiée. L'analyse reste très proche du bruit de fond. Le critère de rejet appliqué au signal réel dans la Partie 5.2.1.4 tend à préserver au plus 5% des valeurs des ions (approximativement 650 points, soit 50 à 100 composés au maximum). Parce que l'inégalité de Bienaymé-Tchebychev est une inégalité de concentration (dans le cadre de la théorie de la probabilité), cette borne est probablement surestimée et pourrait être réduite de moitié. En des termes plus pratiques, la paramétrisation couplée au programme du LNE peut capturer la signature jusqu'à une centaine de composés dans un chromatogramme.

Dans la suite du manuscrit, nous appliquerons systématiquement la comparaison non supervisée d'échantillons à des jeux de données de taille variable (de quelques dizaines à plusieurs centaines de chromatogrammes), qui ne représentent qu'une fraction du marché mondial des emballage dans le contexte d'économie circulaire. Il existe certainement des dizaines de milliers d'emballages distinguables à cet instant, un nombre en croissance permanente. Face à ce défi, la comparaison des séquences de caractères non répétés s'impose avec le choix de sommer la totalité des différences pour obtenir une simple distance scalaire pour chaque paire de données. Il n'est en effet pas envisageable pour un jeu de 100 chromatogrammes de comparer manuellement 5050 paires ($\frac{100 \times (100+1)}{2}$) sans compter les autres avantages de ce choix évoqués dans la Partie 5.2.2.5 tel que l'amélioration de la discernabilité. Cependant toutes les distances estimées entre les données n'ont aucune signification physique, cette limitation devra en partie être mitigée dans les cas d'étude des parties suivantes.

Points clefs

Les bases de la méthodologie qui sera appliquée systématiquement et sans supervision ont été posées. Même si une comparaison avec les lettres répétées est intéressante pour mieux désigner les points de différence entre échantillons, le volume de données à traiter ne peut être fait que sur des scalaires sans supervision. Les séquences sont donc compressées et alignées en vue de la comparaison selon les critères de Levenshtein ou en entropie. Ces comparaisons fixent des distances entre échantillons sans réalité physique pour le moment mais permettant un classement ou une cartographie propre au jeu de données considéré.

5.2.3 . Propriétés et limitations des concepts mathématiques appliqués

La méthodologie de comparaison de signaux chromatographique développée s'appuie sur une comparaison indépendante de l'intensité des signaux de masse et de leur largeur. Seuls leur monotonie et leur signe de variation sont conservés sous la forme d'une séquence de lettres. L'indépendance avec l'intensité est un atout pour analyser la présence de composés inconnus indépendamment du procédé d'extraction (avec ou sans étape de concentration). La seule limite est la détectabilité des fragments et leur séparation par la colonne. Parce que tous les signaux sont considérés, il y a toutefois une plus grande sensibilité au bruit qui peut ajouter de la complexité au problème initial de forêt de pics. Le critère de rejet repose sur le comportement statistique du détecteur. Le choix d'un critère conservatif est préféré car un faux signal positif pourra être détecté par son absence dans une répétition. Les répétitions considérées dans ce travail prennent en compte la variabilité de l'extraction, le bruit thermique de la colonne et les erreurs de l'opérateur. L'indépendance avec la largeur des segments se justifie également mathématiquement. L'intensité du signal original est dans le signal transformé portée par sa hauteur et sa largeur (Figure 3.4). Par ailleurs, l'invariance temporelle en plus du taux de compression extrêmement important (Figure 3.6) garantit une certaine robustesse aux variations des temps d'élution (longueur de colonne, variations du programme de température) et fréquence d'acquisition. Les signaux de deux chromatogrammes encodés temporellement sont comparables en quelques secondes pour tous leurs ions. Deux critères de distance sont proposés suivant que l'on s'intéresse à la longueur de l'édition des séquences (critère de Levenshtein) et/ou à l'information en excès (critère entropique). Les deux critères sont extensifs, ils ne mesurent pas exactement la même chose, mais surtout ils correspondent à deux échelles ouvertes pour lesquelles il n'est pas possible de dire quelle distance correspond à un scénario de type apparition de composés de dégradation, matériau recyclé, absence de décontamination. Il est probable que l'échelle varie suivant l'on s'intéresse à une économie linéaire ou circulaire, mais également en fonction du type de polymère, puisqu'ils correspondent à des stratégies différentes de formulation, d'utilisation et de collecte. L'existence de différences entre mesures répétées complique également la différenciation entre échantillons. Elle peut être comprise soit par un manque de sensibilité des critères (à tester), soit au contraire par la très forte sensibilité du codage à compter toute variation aussi infime soit-elle. Ces deux scénarios seront testés dans un cadre plus général sur la base d'un très grand nombre d'échantillons représentatifs des applications et des usages dans une économie linéaire et circulaire. Il s'agira notamment de vérifier si ces distances permettent de développer un espace chimique dans lequel des effets, des matériaux, des pratiques pourraient être différenciés.

5.3 . Application aux polyoléfines vierges et recyclées

5.3.1 . Vue générale

La séparation des résultats par matrice polymère peut paraître artificielle, mais recouvre une certaine réalité chimique et industrielle. Dans la logique d'une économie circulaire, les polyoléfines retournent peu au contact alimentaire alors que les polyesters comme le PET sont très largement recyclés depuis les deux dernières années. Le polypropylène isotactique (PP) est une matrice polyoléfinique dont les chaînes macromoléculaires présentent une distribution de taille et sont très sensibles à la thermo-oxydation. La distribution de ces molécules dépend non seulement de la conduite de la polymérisation, mais également des conditions de stockage et de mise en forme de ces matrices. Les réactions de scission peuvent ainsi conduire à l'apparition de chaînes plus courtes qui présentent des groupes oxygénés (alcools, aldéhydes, cétones, acides) issus de la décomposition thermique des hydroperoxydes. De très nombreuses chaînes courtes de longueur et polarité variables sont attendues avec le PP avec des chromatogrammes comprenant des forêts de pics dans la région d'élution des oligomères (natifs ou oxydés). Afin de prévenir la dégradation du PP, tous les PP sont enrichis en antioxydants phénoliques, sulfites et phosphites capables de ralentir les réactions d'oxydation. Ces additifs sont des molécules réactives pouvant conduire à des produits de réaction variés plus ou moins stables : adduits, dimères, quinones, etc. Il est ainsi raisonnable de penser qu'il existerait une "signature chimique" (ou une "information chimique") qui tracerait l'origine d'un polypropylène (additivation initiale qui contrôle la durée de vie du matériau), sa formulation (additivation en vue d'obtenir des propriétés mécaniques ou de couleur) et sa dégradation. La Figure 5.15 présente la comparaison des chromatogrammes en PP et rPP. La forêt de pics est visible sur les deux échantillons, leur plus grande complexité peut être comparée à celle dans les échantillons de PET de la figure 5.35. L'identification de différences est particulièrement difficile du fait de la complexité initiale des mélanges. Les analyses mathématiques proposées visent non pas à expliquer chacun des pics (ils sont trop nombreux et peu connus), mais à compter les régions différentes à l'échelle optimale de l'analyse en ondelettes.

En gardant en tête qu'à la différence du PET dont tous les grades vierges sont aptes au contact alimentaire, les PP même vierges peuvent avoir une formulation qui les rend impropres à une utilisation pour le contact alimentaire. De nombreux stabilisants (antioxydants, anti-UV, retardateurs de flamme) et colorants utilisés pour les applications en extérieur (e.g. ameublement), construction automobile et bâtiment, dispositifs électriques et électroniques sont exclus de la liste positive du règlement (EU) 10/2011 relatifs aux matériaux plastiques au contact des aliments. La différence de comportement entre le PET et le PP ne se limite pas à la forte additivation du PP mais également aux valeurs très élevées des coefficients de diffusion du PP ainsi qu'à ses utilisations dans des conditions plus extrêmes de température. Les diffusivités des additifs, oligomères et contaminants exogènes sont de 102 à 105 plus élevées dans le PP ([Fang and Vitrac, 2017](#); [Zhu et al., 2019c](#)). Le PP isotactique est en effet un polymère caoutchoutique dans les conditions normales d'utilisation avec une température de transition vitreuse (T_g) proche de 0°C. Sa nature semi-cristalline lui permet une utilisation jusqu'à 130°C pour les aliments emballés nécessitant une stérilisation. A contrario, le PET avec une T_g comprise entre 76°C et 81°C tolère seulement un remplissage à chaud (requis pour les boissons pasteurisées comme le jus d'orange) jusqu'à 72°C dans les conditions industrielles. Ces raisons expliquent la très forte contamination des PP recyclés par rapport aux PET recyclés.

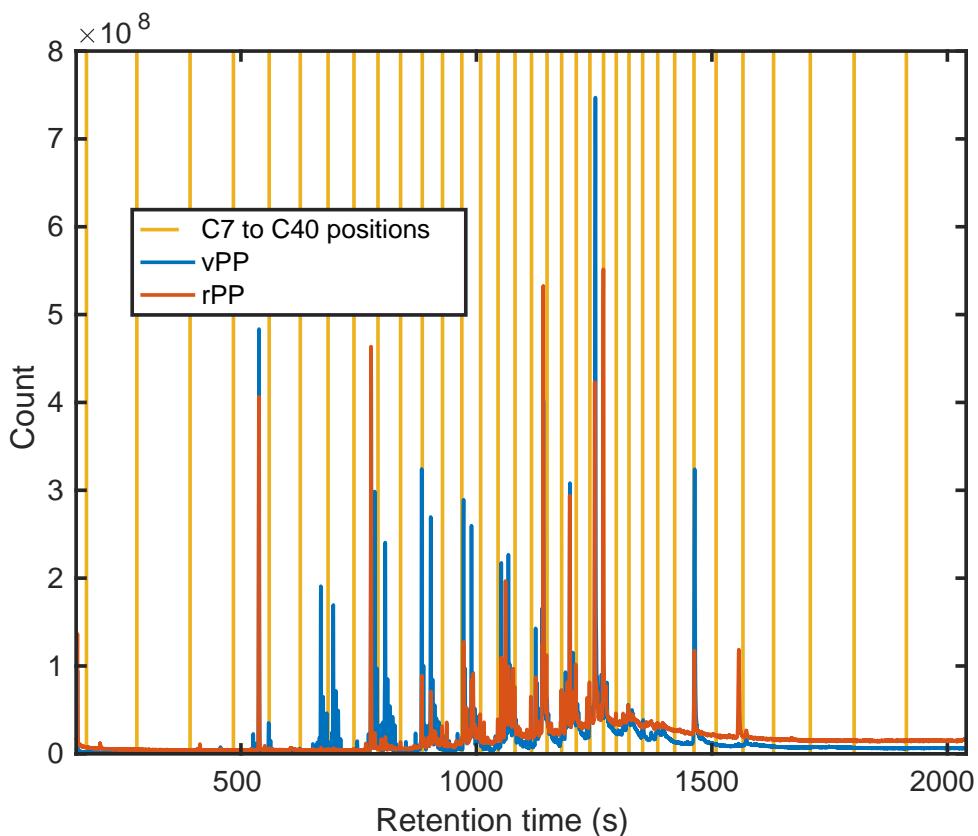


FIGURE 5.15 – Comparaison des TIC issus de l’analyse d’extrait d’un vPP et d’un rPP en GC-MS. Le signal associés aux oligomères dans le rPP sont d’intensité plus faible, les forêts de pics des chaînes les plus courtes disparaissent totalement de la gauche du chromatogramme. La base de la forêt de pic du rPP comporte moins de vallées entre les amas de pics, signalant une plus grande variabilité des espèces par rapport au vPP. La position des alcanes linéaires permet d’estimer la taille des oligomères, les alcanes ramifiés ou oxydés sortant toujours plus tôt que l’équivalent linéaire.

La Partie 5.3.2 présentent une application des méthodes analysées dans la Partie 5.2 sur la reconstruction d’effets industriels (procédé, formulation) susceptibles d’afficher la signature chimique. Cette étude réalisée avec une combinaison de facteurs contrôlés à l’échelle industrielle (conditions reproductibles, lots homogènes) a permis de préciser les capacités de discrimination soit via la distance de Levenshtein entre échantillons deux à deux, soit via une estimation de l’entropie mutuelle entre échantillons. En terme thermodynamique, l’information chimique est ici déterminée par la dynamique endogène du matériau : les composés sont issus de réactions internes sous l’action de la température et de l’oxygène atmosphérique. D’après la chimie du PP, nous nous attendons donc à des variations d’information mineures dans des chromatogrammes déjà très riches.

Il est important pour le lecteur de remarquer qu’aucune orientation ou priorisation de composés n’est donnée *a priori* à la méthode, l’ensemble de l’information est considérée. La capacité à discriminer suivant la qualité de l’information est proposée en comparant le pouvoir de discrimination et de classification des chromatogrammes sur des extraits liquides (composés volatils et semi-volatils de MZ entre 40 et 650) et sur des analyses en espace de tête réalisées par ther-

modésorption (composés volatils de forte concentration de MZ entre 34 et 300). Les réactions d'oxydation et de scission affectent toutes les masses moléculaires. Parce que la méthode d'espace de tête considère un spectre plus réduit de substances et que ces dernières peuvent également être perdues lors du stockage des échantillons, elle est donc censée offrir une information chimique moins exhaustive de l'avancement des réactions au sein du matériau. Parce qu'il n'y a pas d'étape de concentration, la méthode introduit une certaine censure dans l'information disponible du fait de limites de détectabilité plus hautes que pour l'extraction liquide suivie d'une étape de reconcentration.

La Partie 5.3.3 teste la robustesse de l'analyse précédente sur des échantillons complètement indépendants de PP, où une information exogène est rajoutée au matériau PP lui-même vieilli du fait d'une étape de recyclage. L'histoire du PP est ici bien moins contrôlée, car issue de la collecte de déchets ménagers, d'un tri, d'une étape de décontamination et d'une étape de granulation après passage à l'état fondu. La reconstruction d'une traçabilité sur la base d'échantillons pris à deux à deux est soumis à plusieurs difficultés expérimentales. Le saignement de la colonne qui peut conduire à un bruit de Poisson résiduel malgré le traitement numérique appliqué et décrit dans la Partie 5.1. La dérive liée au vieillissement progressif de la colonne chromatographique ou son remplacement est à considérer. Les mesures ont été réalisées avec deux appareils, un pour chaque méthode d'extraction, sur une période continue de 5 mois. Pour l'extraction liquide, l'appareil a réalisé environ 5000 analyses dont 10% ont été dédiées à cette étude, les autres 90% étant utilisés pour les activités de service du LNE. Pour l'extraction gazeuse, le taux d'utilisation est porté à 15-20%. Afin de quantifier l'information non explicative, le protocole complet (y compris la pesée, l'extraction, l'ajout de l'étalon et la reprise, l'injection et la mesure) a été doublé pour chaque échantillon.

Points clefs

Les polyoléfines sont un exemple de matériaux présentant une signature chimique caractéristique avec une forêt de pics associée à la distribution des oligomères et de leurs composés de dégradation. Les différences entre leurs chromatogrammes apparaissent aussi bien sous la forme d'ajouts que de disparitions de composés.

5.3.2 . Identification des effets contrôlant la qualité de pots et de couvercles en PP

5.3.2.1 . Plan d'échantillonnage

Le plan d'échantillonnage des effets porte sur quatre grades de PP utilisés commercialement pour fabriquer des couvercles et des pots destinés au conditionnement de produits gras et aqueux. Quatre effets ont été étudiés :

- Grade du matériau parmi quatre possibles (1 à 4) ainsi que trois autres grades spéciaux (5 à 7) utilisés en mélange pour apporter une coloration ou une étiquette fusionnée
- Format de l'échantillon (granulés ou mis en forme)
- Ajout ou non d'une coloration, d'une étiquette au produit mis en forme (non applicable aux granulés)
- Température de mise en forme (normale à 220°C ou élevée entre 270°C et 290°C selon l'article)

Le plan d'échantillonnage réalisé est contraint par la production des articles (pot ou couvercle), toutes les combinaisons des effets imbriqués ne sont pas possibles. Le plan est présenté de manière synthétique sous la forme d'un arbre présenté sur la Figure 5.16. Les 23 échantillons représentent autant de combinaisons uniques des effets étudiés. Pour chaque matériau, les granulés représentent le matériau de référence ayant subi le moins de transformations (entropie chimique théoriquement la plus faible).

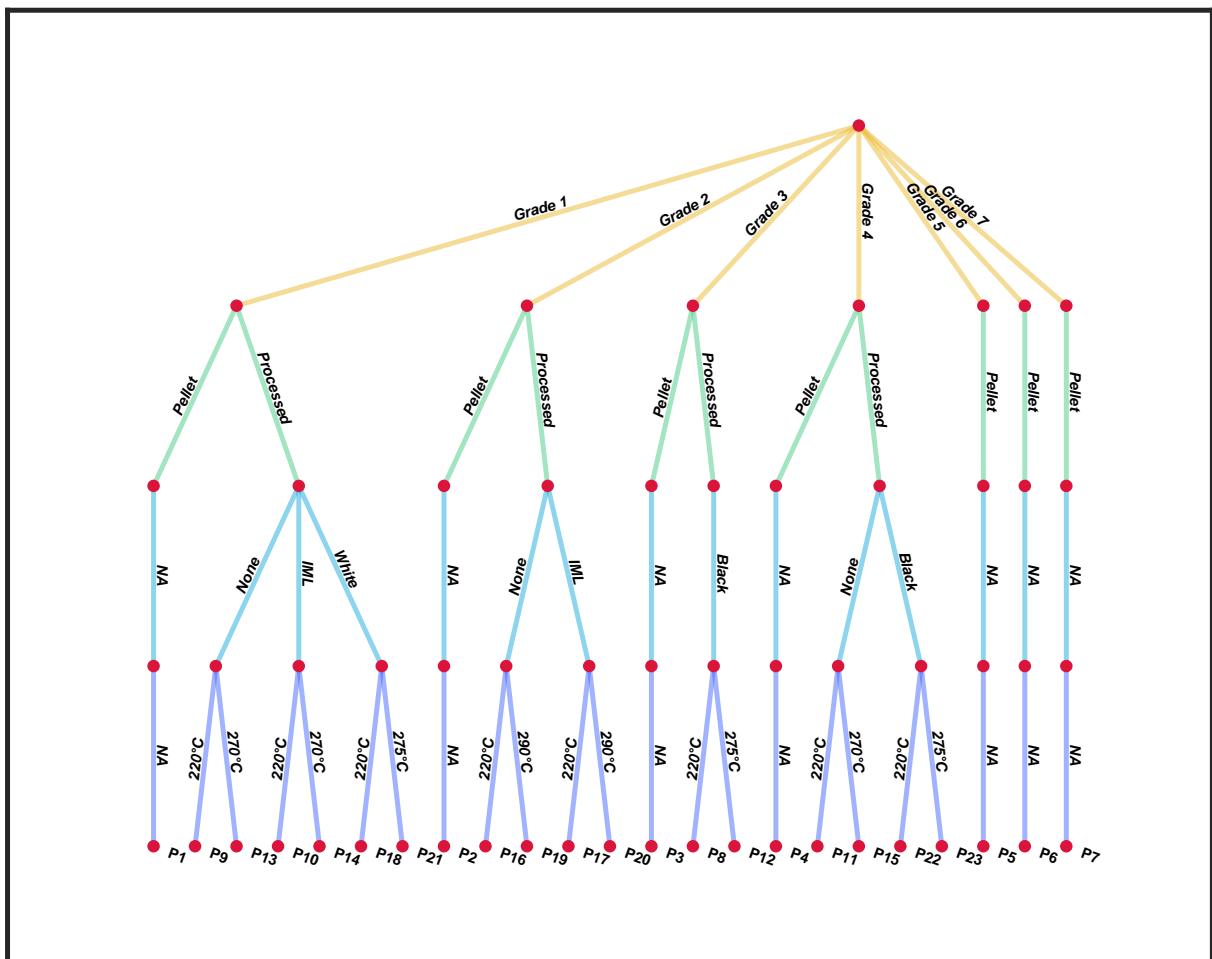


FIGURE 5.16 – Classification des échantillons basée sur les effets a priori. Cette classification n'est pas unique mais représente une des manières possibles de discriminer les échantillons.

Il est important de noter que les échantillons P5, P6 et P7 sont des granulés utilisés uniquement en mélange (ou en ajout) pour obtenir respectivement une étiquette fusionnée (IML), une coloration noire et une coloration blanche. Typiquement, l'échantillon P10 (PP grade 1, mis en forme, additionné d'une IML, à 220°C) pourrait être décrit par l'expression compatible avec l'idée de mélange ($P1 + P5$)₂₂₀.

Points clefs

Le plan d'échantillonnage permet une comparaison entre des granulés (moins transformés) et des articles mis en forme. Quatre résines différentes de grade (grades 1-4) alimentaire sont utilisées quasiment pures comme matrice principale. Trois autres résines (grades 5-7) sont utilisées pour disperser les colorants (noir de carbone, dioxyde de titane) ou former des étiquettes en IML. La mise en forme de leurs combinaisons est testée à deux températures (basse à 220°C, haute entre 270°C et 290°C). Le plan d'échantillonnage imbriqué comprend 23 conditions pour lesquelles deux répétitions ont été réalisées.

5.3.2.2 . Filtrage et orthogonalisation de l'information portée par les combinaisons étudiées

Principes

Les effets étudiés ne peuvent pas être reconstruits directement à partir des distances deux à deux entre échantillons. Ils requièrent un traitement qui permettant de séparer les effets notamment par agrégation des données associées aux mêmes combinaisons des effets. Le passage de distances deux à deux à un espace euclidien de même métrique a été réalisée à l'aide d'une analyse en coordonnées principales (PCoA). Cette technique reconstruit une topologie des échantillons à l'aide d'une décomposition des distances sur les vecteurs propres de la matrice de distance associés aux plus grandes valeurs propres. Les dimensions non-utilisées constituent un sous-espace (possiblement de grande dimension) pour lequel tous les échantillons peuvent être considérés similaires. Cet espace code le bruit et les biais non-explicables par les effets étudiés. L'utilisation d'un nombre de dimensions plus réduit que le nombre d'échantillons réduit la distance entre échantillons, mais permet d'obtenir un filtrage efficace. Dans les cas favorables, les coordonnées principales ou leurs combinaisons peuvent encoder les effets que l'on cherche à étudier. Afin d'étudier la pertinence de cette analyse en coordonnées principales, la pertinence de chaque coordonnée principale été également étudiée en réalisant une analyse de la variance (variance de la distance le long de cette coordonnée) multivariée (en prenant un ou plusieurs effets simultanément).

Décomposition de la distance totale

Les valeurs propres associées à chacune des dimensions générées par la PCoA sont un estimateur de la distance exprimée dans la même dimension. La Figure 5.17 présente le carré de distance à la moyenne (MSD) du cumul des dimensions (classées par ordre décroissant d'importance) en pourcentage du carré de la MSD totale pour les deux critères de comparaison. Pour la distance de Levenshtein nous pouvons observer que les 6 premières dimensions ne comptent que pour 34% de la MSD totale. La pente au-delà de la 6ème dimension est approximativement constante et chaque dimension exprime environ 1% à 2% de la MSDT indiquant qu'elles contiennent majoritairement du bruit. Ne pas considérer ces dimensions revient à ignorer les deux tiers de la MSD totale liée à la distance de Levenshtein et 14 dimensions sont nécessaires pour comptabiliser la moitié de la MSD totale. Le constat est similaire pour la distance en entropie, la pente de la courbe devient quasiment constante à partir de la 5ème dimension, les quatre premières ne comptant que pour 20% de la variance totale de la distance. Toutes les dimensions au-delà de la 5ème comptent chacune pour 2% de la variance totale, le bruit serait alors responsable de 80% de la variance du critère de distance en entropie. Dans ce cas il faut considérer 18 dimensions pour atteindre la moitié de la MSD totale.

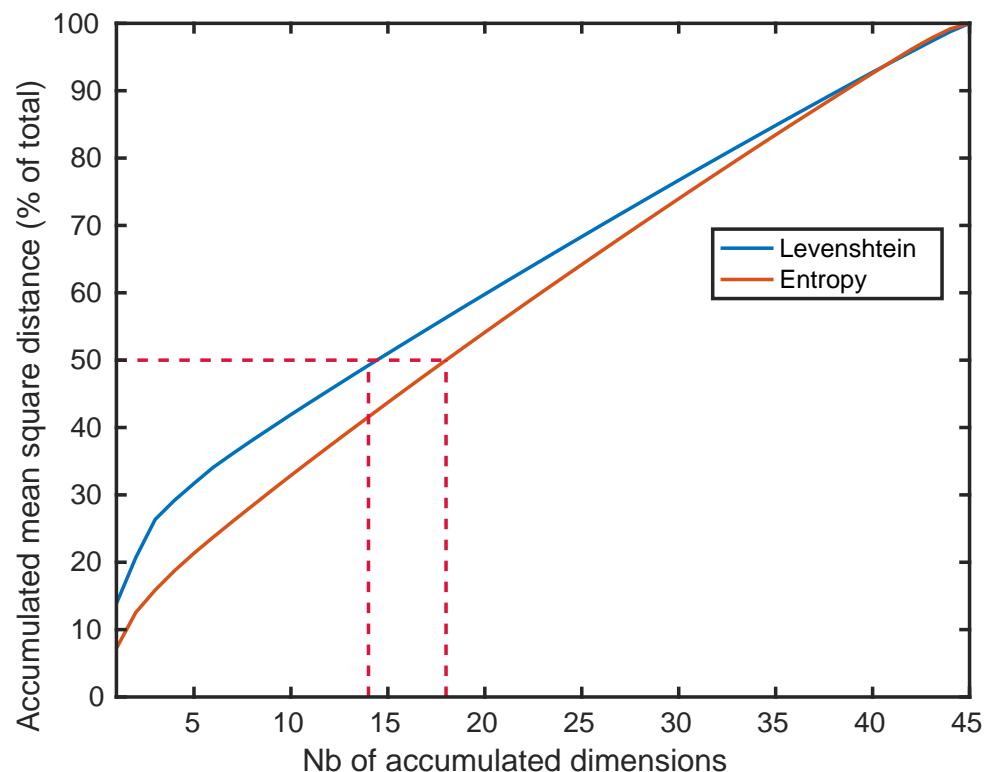


FIGURE 5.17 – MSD du cumul des dimensions en pourcentage du MSD total pour les articles en PP selon les critères de Levenshtein et entropique.

Identification des dimensions significatives

Nous avons pu observer que la distance entre les répétitions de l'analyse d'un même échantillon n'était jamais nulle, indiquant la présence d'un biais systématique et non explicable du point de vue de l'échantillon, mais lié aux propriétés physiques de l'analyse et la méthode numérique appliquée. Ce biais crée une incertitude sur la position réelle de l'échantillon dans nos espaces issus des deux critères de distances et réduit notre capacité à discerner deux échantillons. Nous proposons pour estimer ce biais de distance de considérer le 95ème centile des distances inter répétitions du jeu de données considéré. Le critère ainsi proposé est peu susceptible aux valeurs extrêmes tout en assurant une borne supérieure fiable du biais de distance. La position des échantillons dans les espaces engendrés par les distances de Levenshtein et entropique est estimée par l'isobarycentre de l'ensemble de leurs répétitions (2 par échantillon dans notre cas). Nous considérerons que la distance entre les échantillons correspond à la distance entre leurs positions moyennes respectives. Nous considérons que deux échantillons sont discernables à condition que la distance qui les sépare soit supérieure à l'incertitude entre répétitions. La Figure 5.18 est une évaluation de la capacité à discerner des échantillons en fonction de la dimension de l'espace observée. Les courbes en trait plein présentent le centile 0.10, la médiane et le maximum des distances entre échantillons ainsi que l'estimateur de l'incertitude en traits interrompus.

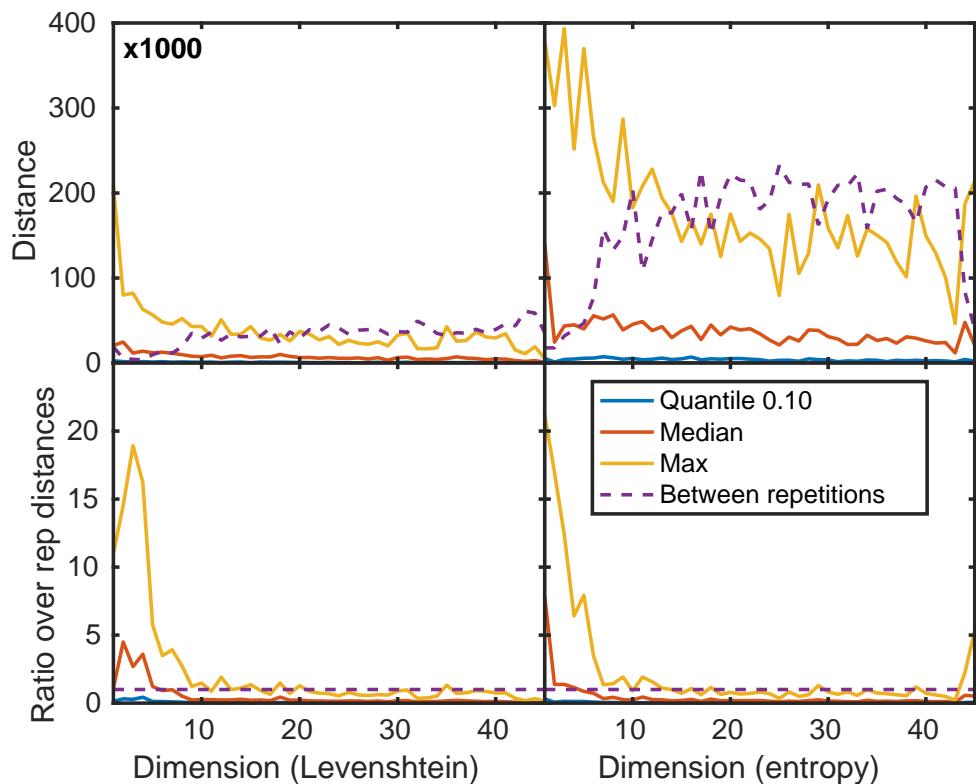


FIGURE 5.18 – Comparaison des distances inter échantillons (10ème centile, médiane et maximum) avec la distance inter répétitions (95ème centile). En haut les distances nettes, en bas le rapport entre distances inter échantillons et inter répétitions.

Il existe au moins un échantillon distant des autres si le critère du maximum dépasse la distance de biais. Le critère de la médiane indique seulement qu'au moins la moitié des distances entre tous les échantillons sont supérieures au biais. Le nombre d'échantillons discernables dépend du nombre de paquets d'échantillons dans l'espace considéré. Enfin le critère du 10ème centile inférieur au biais nous assure qu'au moins une partie des échantillons ne sont pas discernables. Le Tableau 5.7 recense les dimensions d'intérêt selon les critères du maximum et de la médiane pour les distances de Levenshtein et entropique.

TABLEAU 5.7 – Dimensions d'intérêt selon les critères du maximum et de la médiane pour les distances de Levenshtein et entropique appliquées au suivi d'articles en PP analysés en GC-MS.

Critère	Levenshtein	Entropie
Médiane	1 à 5	1 à 4
Maximum	1 à 10, 12, 14, 15, 18, 20, 35	1 à 9, 11 à 13, 16, 18, 29, 39, 44, 45

Les dimensions les plus exploitables sont celles dont le plus petit centile supérieur à l'incertitude est bas. La Figure 5.19 présente le plus petit centile significatif pour chaque dimension. La valeur du premier centile significatif Q indique qu'une proportion Q des distances entre échantillons est inférieure à l'incertitude. Plus Q est faible plus les échantillons sont discernables dans cette dimension. L'absence de point indique qu'aucun échantillon n'est discernable dans cette dimension. Il faut noter que le centile ne donne pas le nombre d'échantillons discernables, seulement la proportion de distances au-dessus ou en dessous de l'incertitude. Le nombre d'échantillons discernables dépend du nombre de groupes formés par l'agglomération des échantillons dans l'espace.

Prenons l'exemple d'un jeu d'échantillons de taille n découpé en deux groupes de tailles respectives k et $n - k$. Il y a alors $\frac{n(n-1)}{2}$ distances uniques (triangle supérieur privé de la diagonale de la matrice de distances paire à paire). Considérons le plus petit centile significatif $Q = 0.5$, nous avons alors $1 - Q = 0.5$ des distances plus grandes que l'incertitude correspondant aux distances de chaque échantillon du premier groupe à chaque échantillon du second groupe au nombre de $k(n - k)$. Nous obtenons ainsi l'Équation 5.11 qui est un polynôme de degré 2 en k à résoudre.

$$k(n - k) = \frac{1}{2} \frac{n(n - 1)}{2} \iff k^2 - nk + \frac{n(n - 1)}{4} = 0 \quad (5.11)$$

Nous avons alors le discriminant et les solutions suivants :

$$\Delta = n^2 - n(n - 1) = n > 0 \quad (5.12)$$

$$k = \frac{n \pm \sqrt{\Delta}}{2} = \frac{n \pm \sqrt{n}}{2} \quad (5.13)$$

Cette analyse sur la médiane et pour l'hypothèse de deux groupes donnerait alors dans notre cas à 23 échantillons deux populations de 14 et 9 individus. Bien évidemment le raisonnement est simpliste et demanderait à être étendu et raffiné sur les points suivants :

- Le centile peut varier de 0 (tous les individus sont discernables) à 1 (un seul groupe homogène du point de vue de l'incertitude).
- Le nombre de groupes pouvant prendre des valeurs entre 1 (équivalent au centile 1) et n le nombre d'individus étudiés (équivalent au centile 0).
- Le cas d'étude considère que les points n'appartiennent qu'à un seul groupe dans lequel tous les individus se trouvent à une distance inférieure à l'incertitude, cependant il existe le cas où deux points distants peuvent en voir un troisième commun.

L'analyse de la Figure 5.19 montre que peu de dimensions permettent de bien différencier les échantillons. Seules 5 dimensions pour le critère de Levenshtein et 4 dimensions pour le critère entropique assurent qu'au plus de la moitié des distances sont inférieures à l'incertitude et permettent une bonne discrimination. Nous prendrons finalement quelques dimensions supplémentaires en sélectionnant celles dont le premier centile significatif est inférieur à 0.60 (au plus 40% des distances sont inférieures à l'incertitude), car l'ajout de une ou deux dimensions n'augmente pas significativement le coût de l'analyse. Nous traiterons alors l'espace issu de la distance de Levenshtein sur 7 dimensions et celui généré par l'entropie sur dimensions.

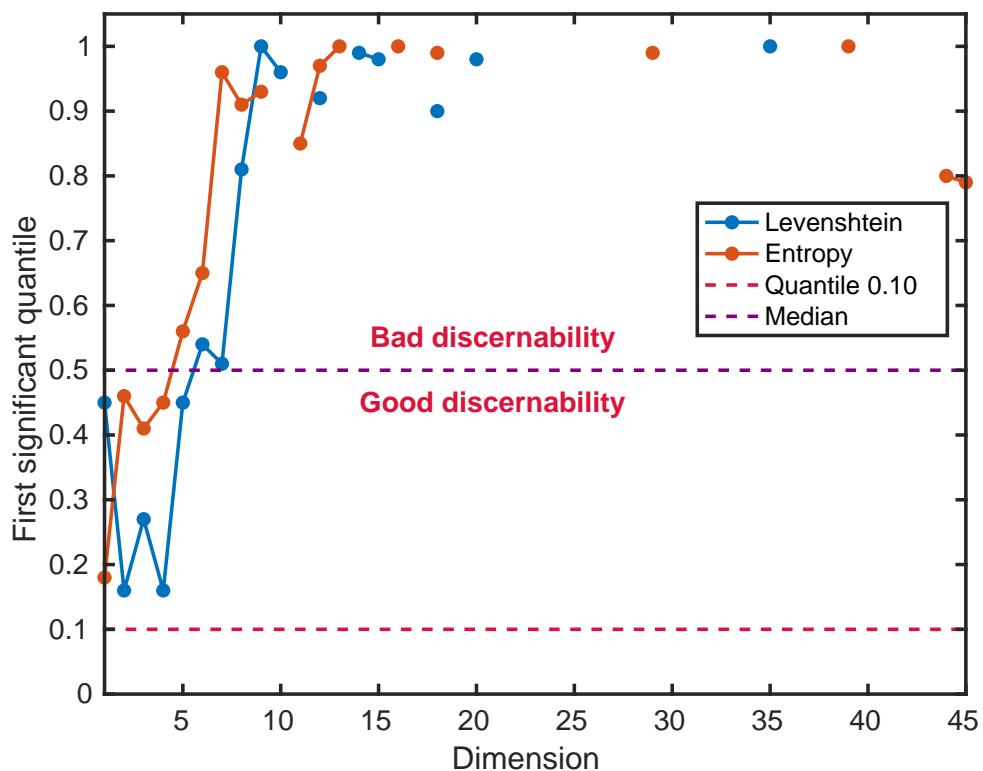


FIGURE 5.19 – Premier centile significatif par dimension pour les distances de Levenshtein et entropique, l'absence de point dans une dimension indique que les échantillons n'y sont pas discernables.

Appariement des effets et des dimensions

Les études précédentes nous ont permis de réduire la taille des espaces à traiter. Nous avons cependant besoin de relier les effets *a priori* de la Figure 5.16 à ces dimensions permettant de séparer en aveugle les échantillons pour espérer répondre à nos objectifs.

Nous avons pour cela utilisé une analyse de variance (ANOVA) à quatre facteurs sans interaction sur nos sous-espaces associés à chaque dimension dont le modèle linéaire général est défini par l'Équation 5.14 :

$$C_{s,d} = T_{s,d} + A_{s,d} + F_{s,d} + M_{s,d} + \epsilon_{s,d} \quad (5.14)$$

où s est un échantillon, d est la dimension considérée, $C_{s,d}$ la coordonnée de s dans d , $T_{s,d}$, $A_{s,d}$, $F_{s,d}$ et $M_{s,d}$ sont respectivement les effets de la température (normale ou haute), de l'ajout (colorant, IML ou aucun), de la mise en forme (granulés ou article) et du grade de matériau (normaux 1, 2, 3, 4 et destinés aux ajouts 5, 6, 7).

La Figure 5.20 présente les résultats de l'ANOVA appliquée aux espaces générés par les deux distances à la fois en effets isolés (4-way) et en effets composés (nested). Notons que l'effet F (mise en forme) n'est pas mesurable seul à cause d'un nombre de degrés de liberté nul. L'effet M (matériau) est majoritaire globalement, suivi des effets F et A (ajout) et l'effet T ("Température") est négligeable.

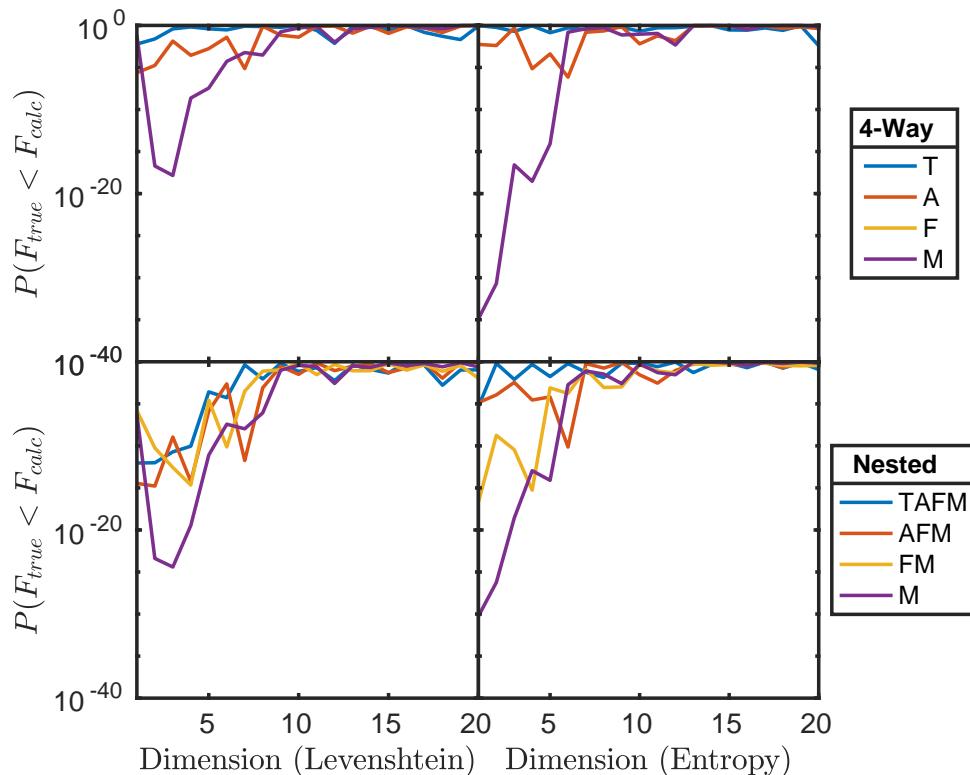


FIGURE 5.20 – ANOVA des effets escomptés (T Température, A Ajout, F Mise en forme, M Matériau) pour le suivi d'articles en polypropylène. La partie haute donne l'ANOVA à 4 effets isolés (effet F non visible à cause du nombre de degrés de liberté nul), la partie basse l'ANOVA avec les effets imbriqués, à gauche l'analyse sur l'espace engendré par la distance de Levenshtein et à droite par la distance entropique.

Points clefs

Les distances des paires entre les 46 mesures chromatographiques développent un espace chimique dont une fraction seulement des dimensions (moins de 7 sur 45) ont un sens physique. L'espace complémentaire est un espace nul qui code le bruit résiduel dans les données. La décomposition en coordonnées principales permet de justifier que la moitié de la variance observée entre échantillons ne signe pas pour des empreintes chimiques explicables. Contrairement à l'intuition, l'effet explicable est toutefois supérieur à quelques pour cent. A ce stade la méthode ne sépare pas si ces effets détectables sont liés à des ajouts (IAS), des apparitions ou contaminations (NIAS) ou des disparitions (volatilisation, dégradation) de composés.

5.3.2.3 . Analyse des effets sur l'information chimique extraite par voie d'analyse par solvant

Nous utilisons l'ANOVA pour extraire les dimensions d'intérêt vis-à-vis de chacun des effets et ainsi recréer un sous-espace maximisant la dispersion explicable des échantillons. Le Tableau 5.8 présente les dimensions retenues pour chacun des effets escomptés à la fois pour la distance de Levenshtein ainsi que pour l'entropie dans la limite des dimensions exploitables retenues précédemment (seuil à 60%).

TABLEAU 5.8 – *Dimensions d'intérêt pour expliquer les effets escomptés du suivi d'articles en polypropylène pour les distances de Levenshtein et en entropie.*

Effet	Levenshtein	Entropie
T Température	/	/
A Ajout	1, 2, 4, 7	4
F Mise en forme	2, 3, 5	1, 4
M Matériau	2, 3, 4	1, 2, 3

L'ANOVA permet d'associer aux dimensions restantes un ou plusieurs des effets escomptés, en choisissant intelligemment quelles dimensions combiner et représenter, les échantillons doivent se placer les uns par rapport aux autres en fonction de leurs caractéristiques.

Les effets escomptés sont dilués dans les sept dimensions retenues pour la distance de Levenshtein. La Figure 5.21 donne deux exemples de projections sur les plans générés par les dimensions 1×2 et 2×4 . Ces deux projections proposent une séparation des échantillons principalement selon les effets M et F sauf pour les grades 1 et 3 complètement confondus dans la projection 1×2 et confondus à l'état de granulés dans la projection 2×4 . Cependant aucun des effets indiqués par un segment de couleur ne conserve une orientation constante dans ces projections, indiquant que les effets sont probablement mélangés dans chacune des dimensions. Nous pouvons aussi noter les points P21 et P17 sont respectivement très éloignés des points P18 et P20 qui ne diffèrent que par le température de mise en forme. Cet écart n'est pas lié à une erreur aléatoire au moment de la préparation des échantillons ou de leur analyse, car les deux répétitions se trouvent bien à proximité l'une de l'autre.

Les effets sont moins dilués dans l'espace généré par la distance en entropie, permettant de les représenter ou non en choisissant convenablement les dimensions de projection. La Figure 5.22 propose deux projections sur les dimensions 1×2 et 1×4 . La projection sur les dimensions 1×2 permet le meilleur éclatement de la population d'échantillons en accord avec leurs MSD respectives de 7% et 5%. Les échantillons sont aussi discriminés selon les effets M et F avec toujours la confusion entre les grades de matériaux 1 et 3. La seconde projection sur les dimensions 1×4 est plus intéressante, la séparation précédente est conservée par la dimension 1 et la dimension 4 permet d'obtenir un alignement des effets F et A pour l'ensemble des échantillons. L'effet F indiqué par une flèche verte est orienté de haut en bas en diagonale pour le passage des granulés aux produits mis en forme. L'effet A indiqué par les flèches en bleu est plus faible. En prenant l'article sans ajout comme référence, les articles équivalents colorés en noir ou en blanc sont déplacés vers le haut tandis que les articles avec IML sont déplacés vers le bas.

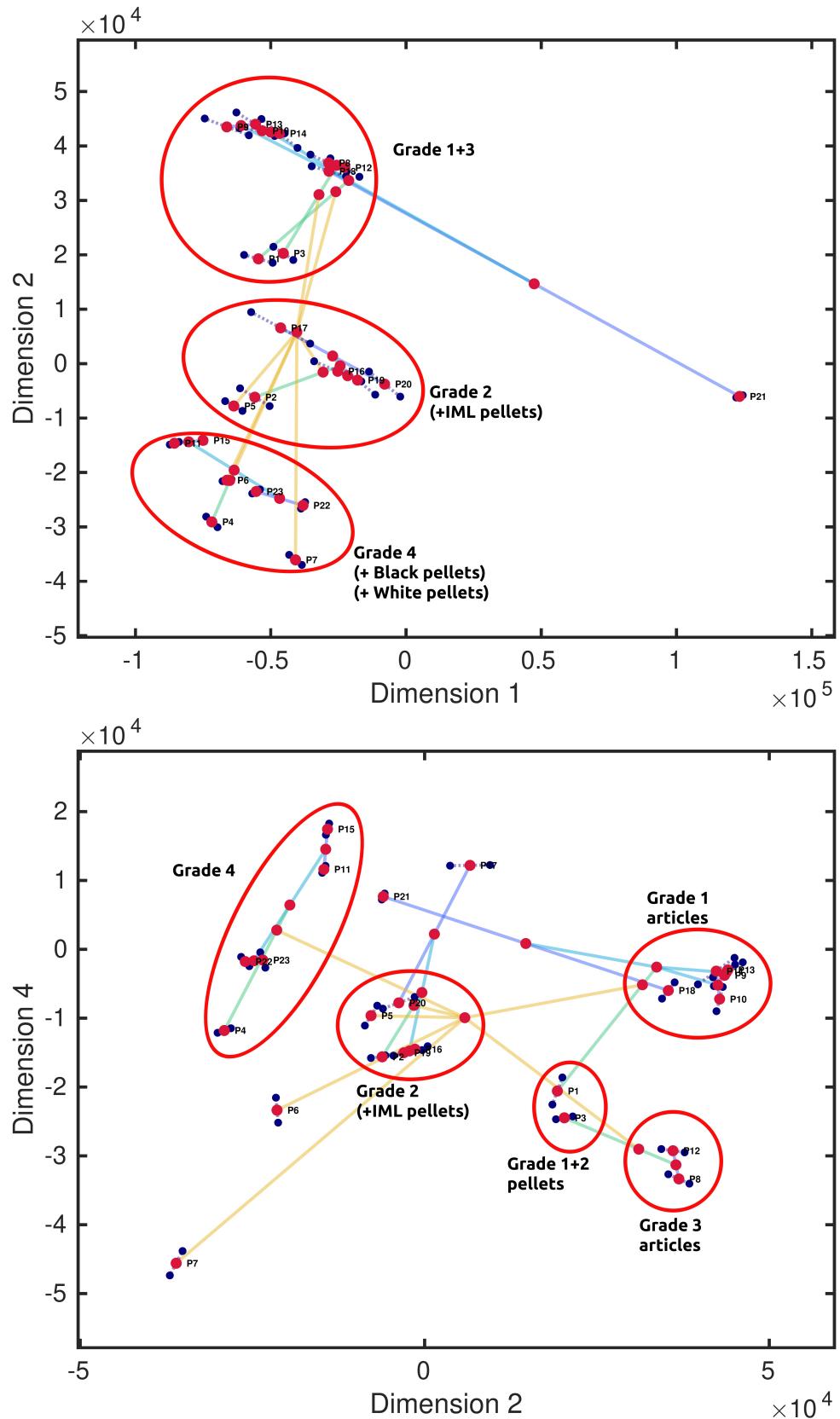


FIGURE 5.21 – Exemples de projections dans l'espace généré par la distance de Levenshtein. En haut la projection 1×2 , avec une séparation par les effets M et F sauf pour les matériaux de grades 1 et 3 confondus. En bas la projection 2×4 séparant aussi selon les effets M et F, mais permettant de séparer les articles mis en forme de grade 1 et 3.

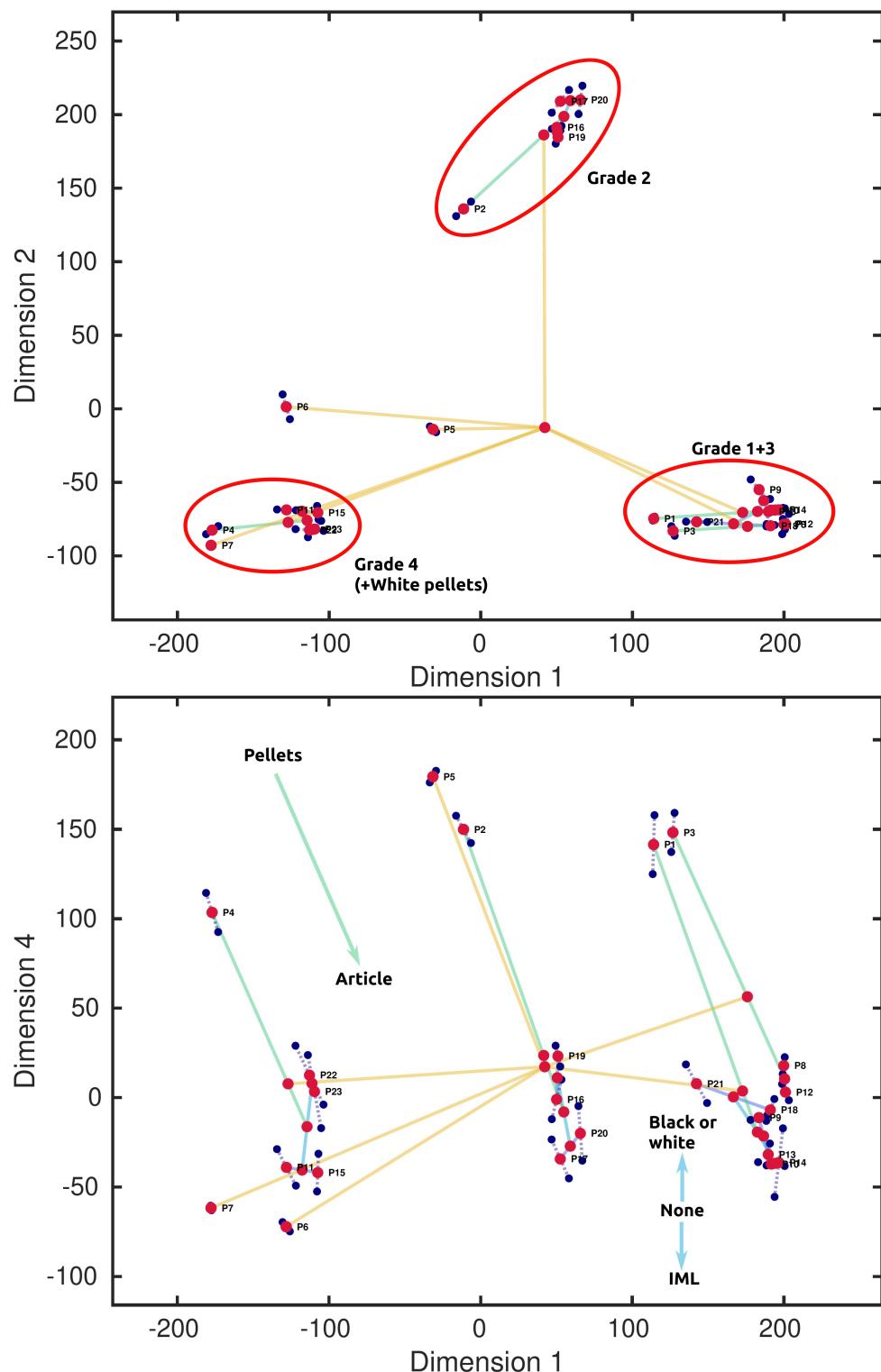


FIGURE 5.22 – Exemples de projections dans l'espace généré par la distance en entropie. En haut la projection 1×2 , avec une séparation par les effets M et F sauf pour les matériaux de grades 1 et 3 confondus. En bas la projection 1×4 mettant en évidence les directions favorisées par les effets F et A . L'effet F indiqué par une flèche en diagonale de haut en bas signe le passage du granulé au produit mis en forme. En prenant comme référence le produit mis en forme sans ajout, l'effet A indiqué par les flèches en bleu se déplace vers le haut pour les colorations en noir ou en blanc, et vers le bas pour l'ajout d'une IML.

Points clefs

L'analyse de la variance imbriquée à quatre facteurs permet de mettre en évidence que la combinaison des effets est détectable avec les dimensions principales retenues. L'effet des grades est le plus visible, suivi de la mise en forme ou non. L'arbre des causes initiales (origine, formulation, mise en forme) n'est pas déformé par la métrique des de la décomposition des distances dans l'espace des échantillons. La causalité apparaît comme un déplacement chimique dans l'espace des dimensions principales : les déplacements sont vectoriels (avec une direction et une norme reproductibles).

5.3.2.4 . Analyse des effets sur l'information chimique extraite par voie d'analyse en phase gazeuse

Les données obtenues par analyse en HS-GC-MS sont traitées de manière analogue aux données GC-MS. La Figure 5.23 présente la MSD cumulée en fonction du nombre de dimensions en pourcentage de la MSD totale des espaces générés par la distance de Levenshtein et la distance en entropie. Pour la distance de Levenshtein, nous pouvons observer une diminution de la MSD par dimension ajoutée à partir de la 6^{ème} dimension, les cinq premières cumulant 63% de la MSD totale. Pour la distance en entropie, la cassure se fait dès la 3^{ème} dimension, les deux premières ne cumulant que 14% de la MSD totale. La distance de Levenshtein semble plus intéressante selon ce critère car l'espace généré concentre les deux tiers de son étendue sur seulement cinq des 45 dimensions. A l'inverse, la distance en entropie a tendance dans ce cas à diluer les distances sur l'ensemble des dimensions.

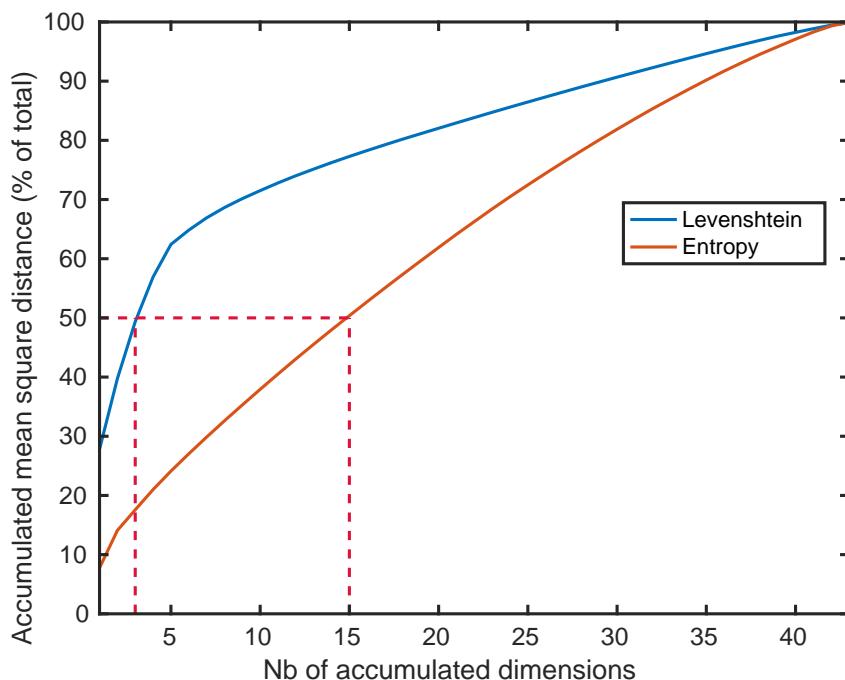


FIGURE 5.23 – MSD associée au cumul des dimensions en pourcentage de la MSD totale pour les distances de Levenshtein et entropique en analyse HS-GC-MS.

L'analyse de la discernabilité des échantillons par rapport à l'incertitude calculée sur la base du centile 0.95 des distances entre répétitions est donnée en Figure 5.24. Nous pouvons observer que la plupart des dimensions ne permettent pas de discriminer les échantillons, seules les dimensions 2, 4 et 8 pour la distance de Levenshtein et les dimensions 1 et 2 pour la distance en entropie permettent de discerner plus de 50% des distances. En choisissant le critère des 60% utilisé dans notre analyse des données GC-MS, nous conserverons les dimensions 2,4, 7 et 8 pour la distance de Levenshtein et seulement les dimensions 1 et 2 pour la distance en entropie.

La Figure 5.25 présente les résultats de l'ANOVA à quatre effets sans interactions et de l'ANOVA avec effets composés d'après les informations disponibles *a priori* sur les échantillons (T Température, A Ajout, F Mise en forme, M Matériau). En ne conservant que les dimensions extraites de l'analyse de discernabilité au centile 0.60 (2, 4, 7, 8 pour la distance de Levenshtein et

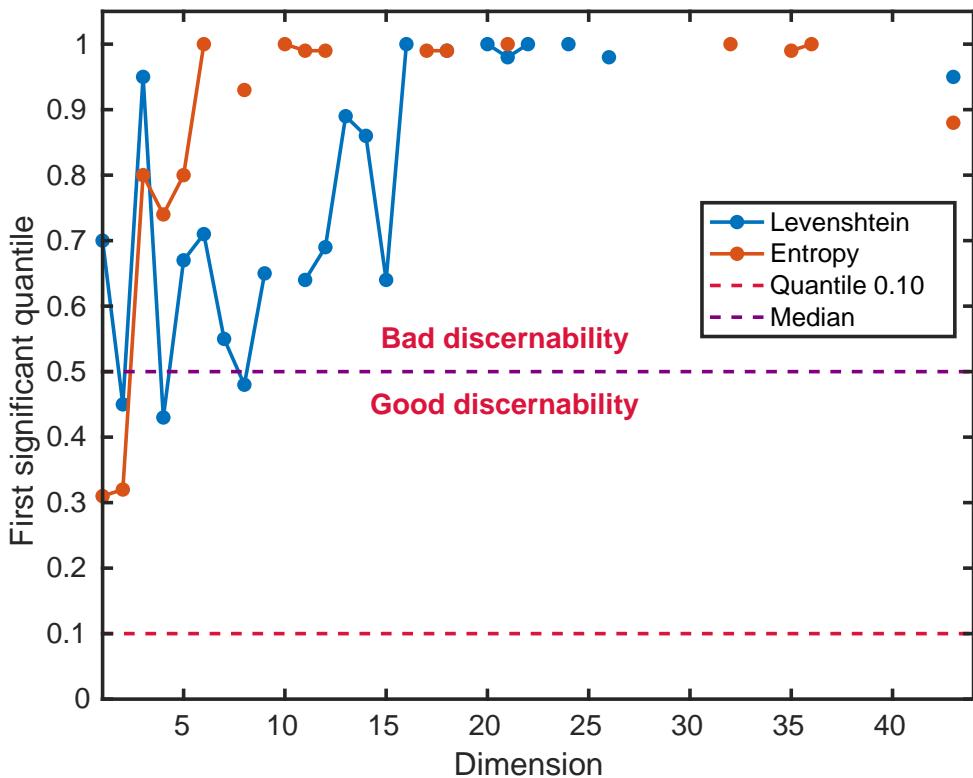


FIGURE 5.24 – Premier centile significatif par dimension pour les distances de Levenshtein et entropique en HS-GC-MS, l’absence de point dans une dimension indique que les échantillons n’y sont pas discernables.

1, 2 pour l’entropie), le Tableau 5.9 compile les dimensions dans lesquelles les effets s’expriment d’après les ANOVA. L’effet M est majoritaire suivi de l’effet F, l’effet A n’est visible que pour la dimension 1 en entropie et l’effet T est négligeable dans tous les cas. Les effets semblent plus mélangés et dilués qu’en analyse GC-MS, il ne devrait donc pas avoir de direction préférentielle dans les différentes projections possibles.

TABLEAU 5.9 – Dimensions d’intérêt pour expliquer les effets escomptés du suivi de production d’articles en polypropylène pour les distances de Levenshtein et en entropie en HS-GC-MS

Effet	Levenshtein	Entropie
T Température	/	/
A Ajout	/	1
F Mise en forme	4, 7, 8	1
M Matériau	2, 4, 7, 8	1, 2

La Figure 5.26 présente un exemple de projection en distance de Levenshtein sur les dimensions 2×4 . Cette cartographie n’a que peu d’intérêt, elle ne sépare que les échantillons de grade 4 des autres non discernables et l’incertitude (segments en pointillés entre deux répétitions représentées par des points bleus) est grande par rapport aux distances entre échantillons. Cela était prévisible avec les premiers centiles significatifs donnés par la Figure 5.24 qui sont à peine inférieurs à 0.50 pour les trois dimensions principales.

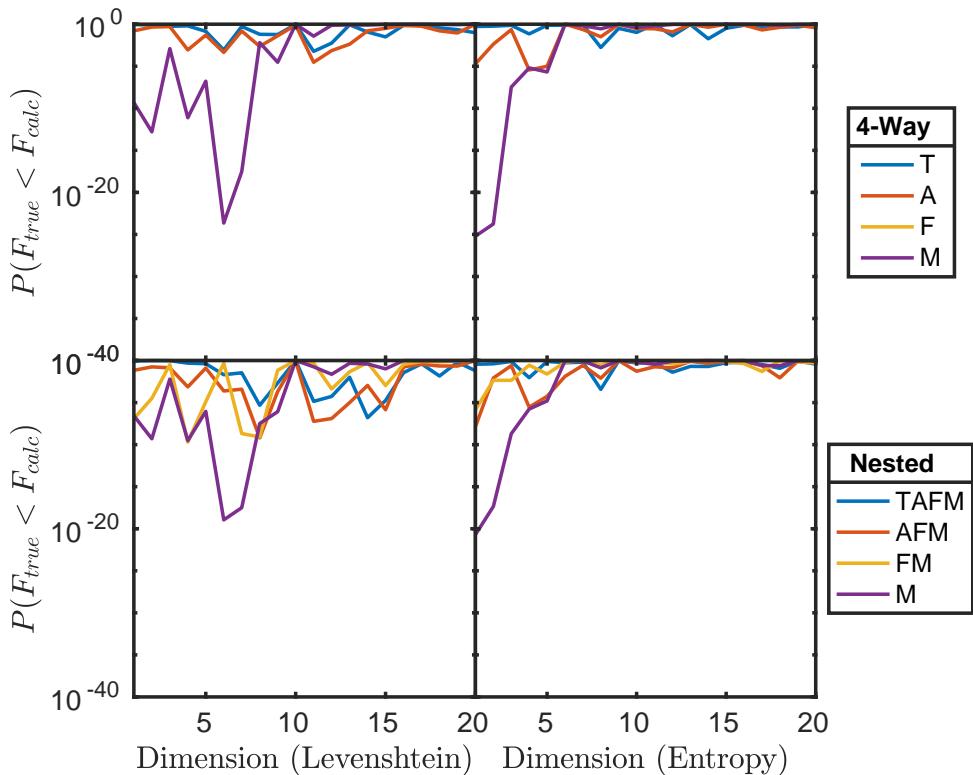


FIGURE 5.25 – ANOVA des effets escomptés (*T* Température, *A* Ajout, *F* Mise en forme, *M* Matériau) pour le suivi d’articles en polypropylène en HS-GC-MS. La partie haute donne l’ANOVA à 4 effets isolés (effet *F* non visible à cause du nombre de degrés de liberté nul), la partie basse l’ANOVA avec les effets imbriqués, à gauche l’analyse sur l’espace engendré par la distance de Levenshtein et à droite par la distance entropique.

L’unique projection d’intérêt d’après nos analyses de l’espace généré par la distance en entropie est celle sur les dimensions 1×2 et est illustrée dans la Figure 5.27. Cette projection permet la séparation selon l’effet *M* des échantillons en grades 2, 4 et du rassemblement des grades 1 et 3. Contrairement à l’analyse en GC-MS, aucune direction n’est associable à un effet en particulier.

Conformément aux limites de l’analyse en HS-GC-MS, la quantité d’information contenue dans les chromatogrammes plus limitée ne permet pas une séparation aussi poussée qu’en GC-MS. En contrepartie, ce type d’analyse ne nécessite pas une longue préparation d’échantillon et pourrait servir de première approximation à la classification des matériaux.

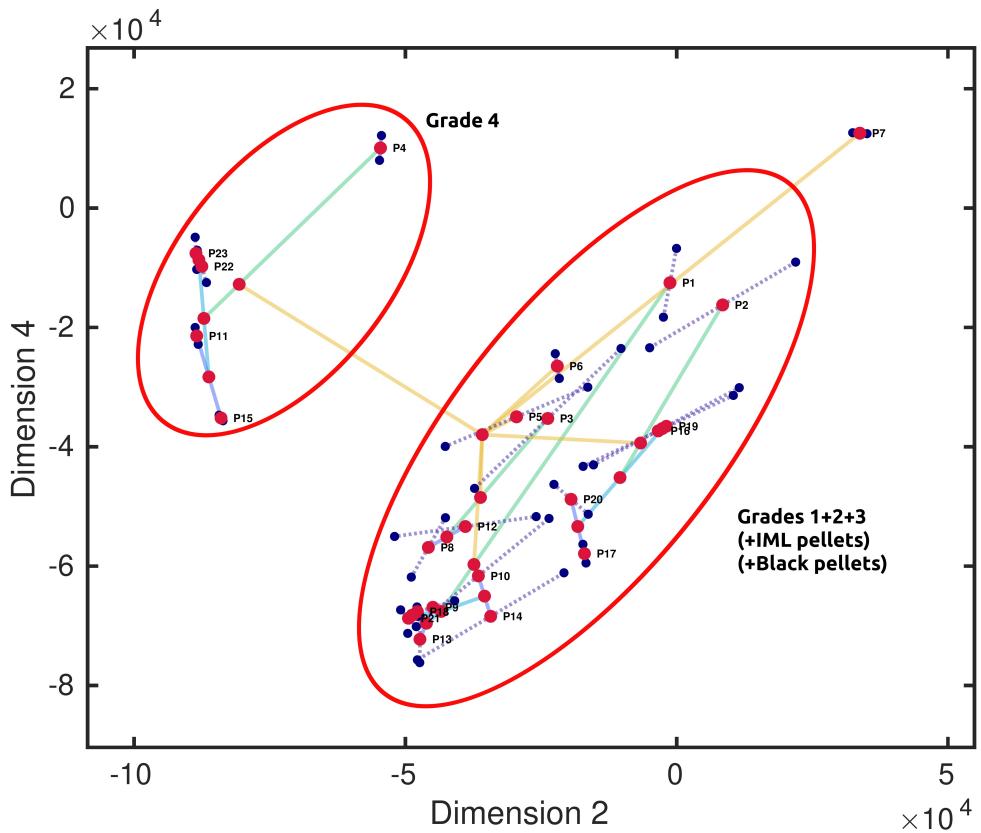


FIGURE 5.26 – Exemple de projection dans l'espace généré par la distance de Levenshtein en HS-GC-MS. La projection (2,4) ne permet que la séparation des échantillons en matériau de grade 4 des autres, l'incertitude est également très grande et est représentée par les traits en pointillés entre deux répétitions (points bleus).

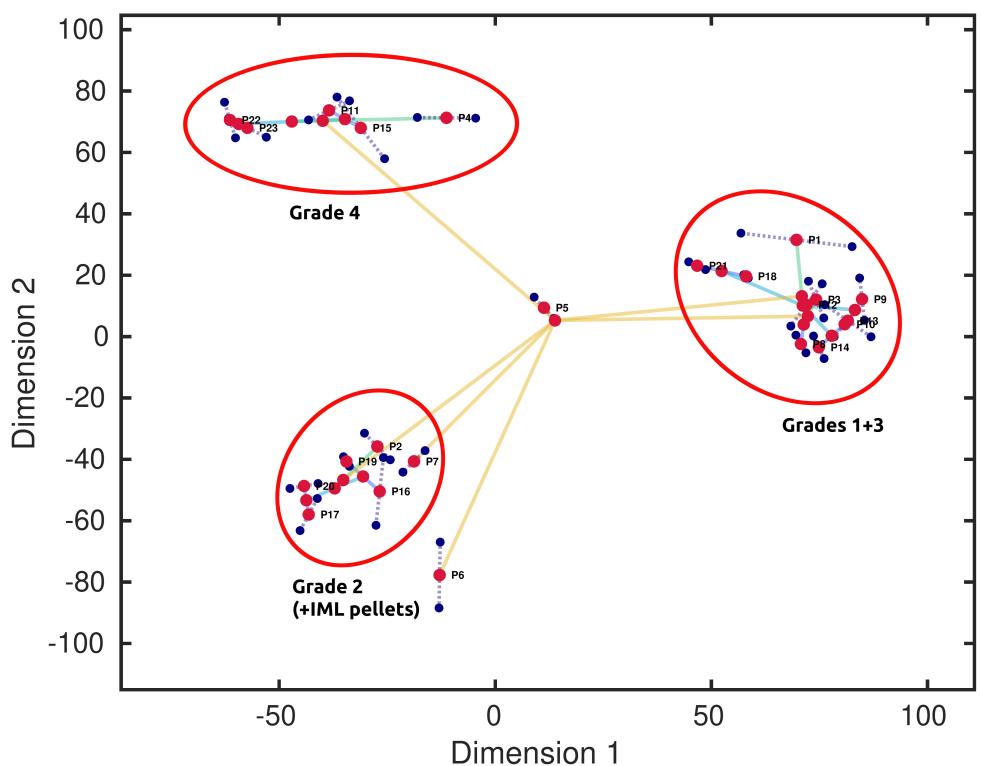


FIGURE 5.27 – Exemple de projection dans l'espace généré par la distance en entropie en HS-GC-MS. La projection (1,2) permet de discriminer les échantillons en matériau de grades 2, 4 et de l'ensemble 1 et 3.

Points clefs

Les données obtenues en HS-GC-MS contiennent moins d'information qu'en GC-MS avec extraction liquide. La méthode ne peut voir que les composés volatils les plus concentrés et cela affecte la capacité à différentier les mêmes échantillons de PP puisqu'ils ne sont comparés que sur une fraction réduite de leur information chimique totale.

5.3.3 . Une première tentative de validation et de généralisation : comparaison de l’information chimique extraite d’échantillons de PP recyclés

5.3.3.1 . Passage d’une cartographie locale (sur site de production) à une cartographie globale dans le cadre de l’économie circulaire

Les vecteurs propres qui permettent de passer d’un espace de distances paire à paire à un espace euclidien dépendent des échantillons considérés et surtout de leur nombre. L’espace développé doit être vu comme un sous-espace convexe dans lequel il n’est pas possible d’imaginer correctement ce qu’il y a en dehors de cet espace. Si nous prenons une analogie, les cartes marines de l’antiquité centrée sur la *Mare Nostrum* ne permettraient pas d’envisager les distances transatlantiques ou autour de l’Afrique. De nombreux facteurs, dont la sphéricité de la terre, ne seraient pas pris en compte. Les vecteurs propres évalués sur une production d’usine ne permettraient pas d’obtenir les distances entre échantillons obtenus dans le cadre d’une économie circulaire. Dans cette partie, nous allons explorer la capacité à reconstruire une carte à grande échelle entre des échantillons de PP très différents, car issus de modes de collecte et de recyclage variés. Dans cette perspective, le PP de référence n’a probablement pas besoin d’être aussi détaillé (effets de température, grade), car nous nous intéressons à des différences informationnelles beaucoup plus grandes apportées par des contaminants exogènes variés et une dégradation plus marquée.

Une carte de l’espace des PP recyclés a été explorée en prenant 5 échantillons recyclés et décontaminés sur le marché nord-américain. Il est important de noter que le marché du PP recyclé est inexistant sur le marché européen. Par analogie avec la Figure 5.16 et pour aider la lecture des résultats, les échantillons sont classés également sous la forme d’un arbre présenté en Figure 5.28. Ils diffèrent par leurs zones et modes de collecte :

- Un échantillon provient d’une collecte indifférenciée aux USA et Canada (G2)
- Un échantillon provient d’une collecte indifférenciée aux USA (G3)
- Trois échantillons (noté G4) provenant d’une collecte aux USA avec un tri sélectif industriel automatisé visant à privilégier l’origine alimentaire des matériaux recyclés.

La référence PP vierge est représentée par un PP atactique vierge de grade commercial (G1) qui n’a subi aucune mise en forme. La cartographie attendue est bien plus grande que celle de la *Mare Nostrum* de la partie 5.3.2, car les échantillons couvrent également des utilisations et des formulations non alimentaires du PP.

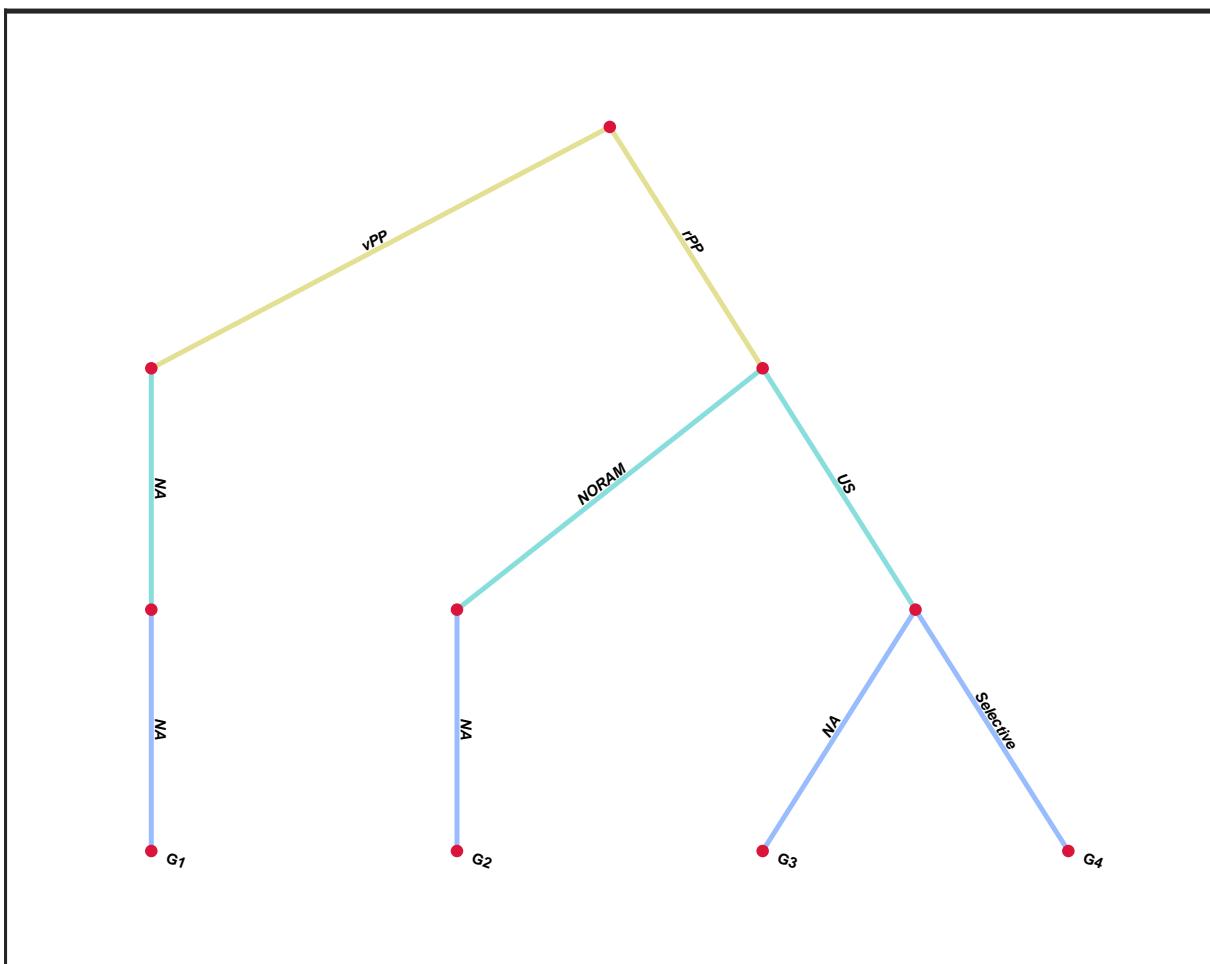


FIGURE 5.28 – Classification des échantillons de polypropylène vierges et recyclés basée sur les effets *a priori*. Cette classification n'est pas unique mais représente une des manières possibles de discriminer les échantillons.

Points clefs

L'espace chimique a été vu initialement comme un espace s'étendant suivant des causes. Ces mêmes causes entraînent des effets dans les conditions contrôlées de l'essai. Mais dans le cas des matériaux recyclés, des causes ou effets non prévisibles s'ajoutent et rendent l'espace de l'information chimique peu compréhensible du point de vue des effets attendus.

5.3.3.2 . Calibration de la cartographie des PP recyclés

L'analyse réalisée est en tout point semblable que celle réalisée dans la Partie 5.3.2.2. La Figure 5.28 présente la distance quadratique totale expliquée par les dimensions cumulées de l'analyse en coordonnées principales (PCoA) pour les deux critères retenus. Le petit nombre d'échantillons ne permet pas d'identifier clairement les deux régimes : l'un où la distance augmente rapidement avec le nombre de dimensions et un autre où la distance totale augmente linéairement avec la dimension en accord avec une orthogonalisation régulière du bruit sur toutes les dimensions de l'espace nul. Un calcul des pentes montre que la zone de transition apparaît après deux ou trois coordonnées principales. Le critère de Levenshtein produit moins de bruit parce que les distances sont plus grandes pour le même nombre de dimension. L'identification des centiles significatifs à l'aide de la Figure 5.30 confirme l'analyse précédente. Les points inférieurs à la médiane identifient des dimensions qui permettent une séparabilité des distances supérieure à un choix aléatoire. Les deux premières dimensions offrent un risque de mauvaise séparabilité compris entre 30% et 40%. La combinaison de deux dimensions réduit ce risque à moins de 10%.

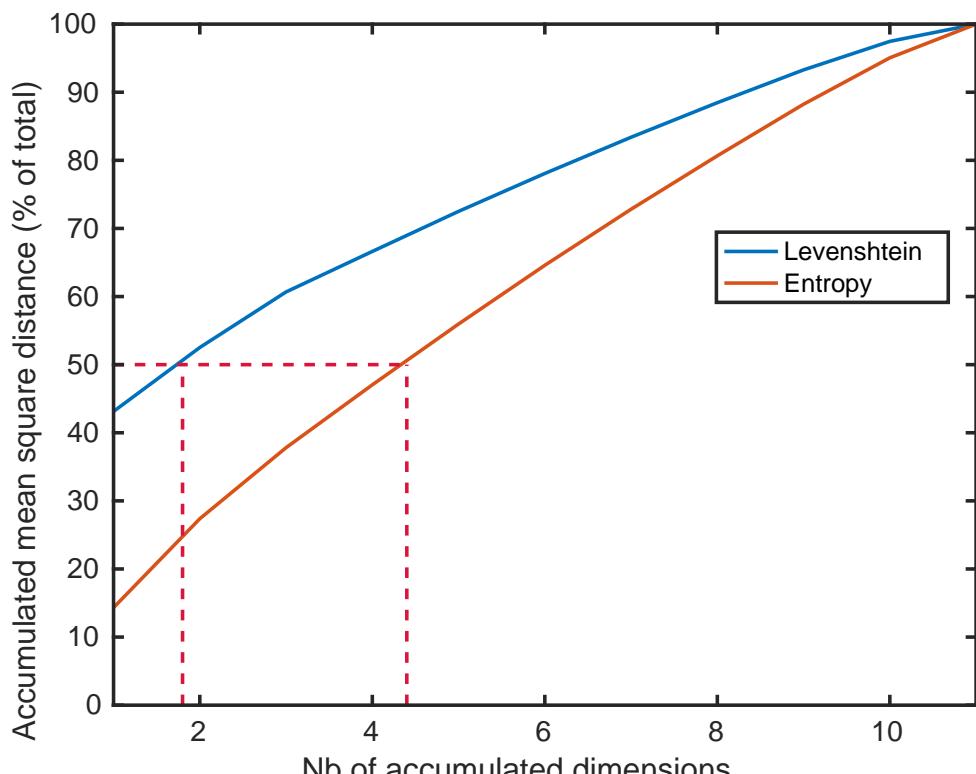


FIGURE 5.29 – MSD du cumul des dimensions pour les polypropylènes vierges et recyclés en pourcentage de la MSD totale pour les distances de Levenshtein et entropique.

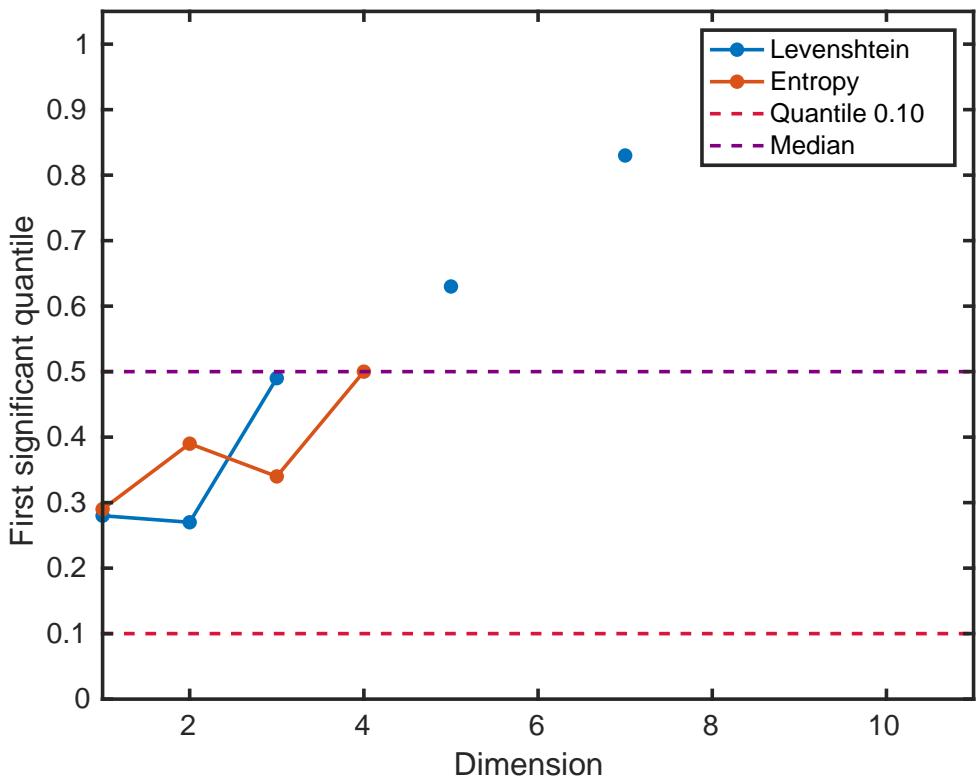


FIGURE 5.30 – Premiers centiles significatifs pour l'espace des PP recyclés.

Points clefs

L'espace chimique est reconstruit de la même manière que dans la Partie 5.3.2.3, mais sans éléments physiques tangibles autres qu'une qualification de la zone géographique de collecte et du type de tri. La cause est vue est ici comme une distance à la matière vierge correspondante.

5.3.3.3 . Reconstruction de la traçabilité et cartographie des PP recyclés

La capacité des distances à reconstruire la nature recyclée ou non (effet R), l'origine (effet O) ou le tri (S) est testée sur la Figure 5.31 en réalisant soit une analyse à trois effets imbriquée (3 effets testables) ou non (2 effets testables). L'analyse montre que les trois premières coordonnées sont capables de séparer les itinéraires d'économie circulaire : SOR, OR et R. En d'autres termes, elles identifient la matière recyclée sans ambiguïté quelque soit le critère retenu. Elles identifient la combinaison des origines et du tri.

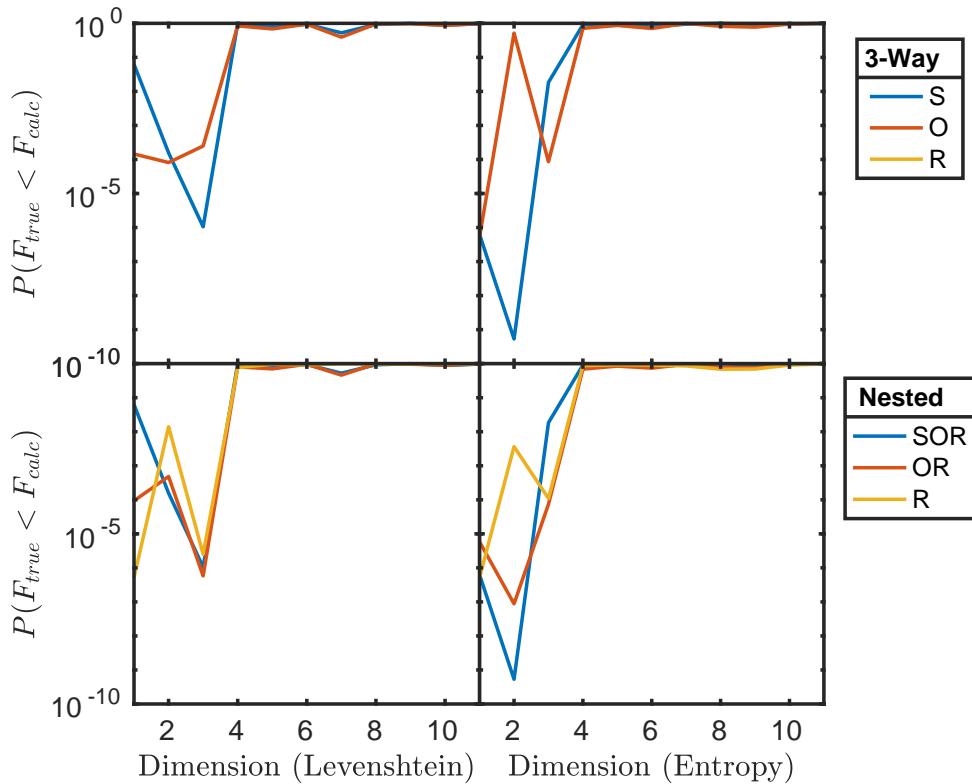


FIGURE 5.31 – ANOVA à trois effets (imbriqués et isolés) sur les PP recyclés. S correspond au tri (aucun ou sélectif), O à l'origine du gisement (non applicable, États-Unis ou Amérique du Nord) et R au caractère recyclé (vierge ou recyclé mécaniquement).

Les cartographies pour le critère de Levenshtein correspondantes suivant les coordonnées 1×2 et 1×3 sont représentées sur la Figure 5.32. Elles montrent des échantillons recyclés très éloignés de l'échantillon vierge. Les distances sont typiquement 2 à 3 fois plus grandes que celles observées entre un granulé et sa mise en forme choisie comme distance de référence (Figure 5.21). Nous confirmons ainsi que le domaine exploré est distinct du domaine exploré dans la Partie 5.3.2.2. La dimension 1 tend à séparer les recyclés de la matière vierge avec une dimension 2 qui sépare les matières recyclées entre elles. Les distances entre matières vierges et recyclées sont équivalentes entre les positions les plus extrêmes des recyclés. L'espace développé par les recyclés eux-mêmes est probablement plus grand, mais son échantillonnage est limité par le nombre d'observations disponibles. Les incertitudes représentent grossièrement 10% des distances maximales.

Les deux critères couplés à un filtrage *ad hoc* donnent des conclusions similaires avec des différences qui sont soulignées ci-après. Les cartes présentent également les arbres de classification dans cet espace. Le critère de discrimination des échantillons préserve la classification *a priori* des échantillons si la topologie des branchements n'est pas affectée (pas de croisements). La distance filtrée de Levenshtein permet de garder une forme stable alors que la distance construite sur

l'information mutuelle filtrée entre les échantillons tend à replier le PP vierge (G1) vers le PP recyclé (G3) sur les projections 2×3 de la Figure 5.33.

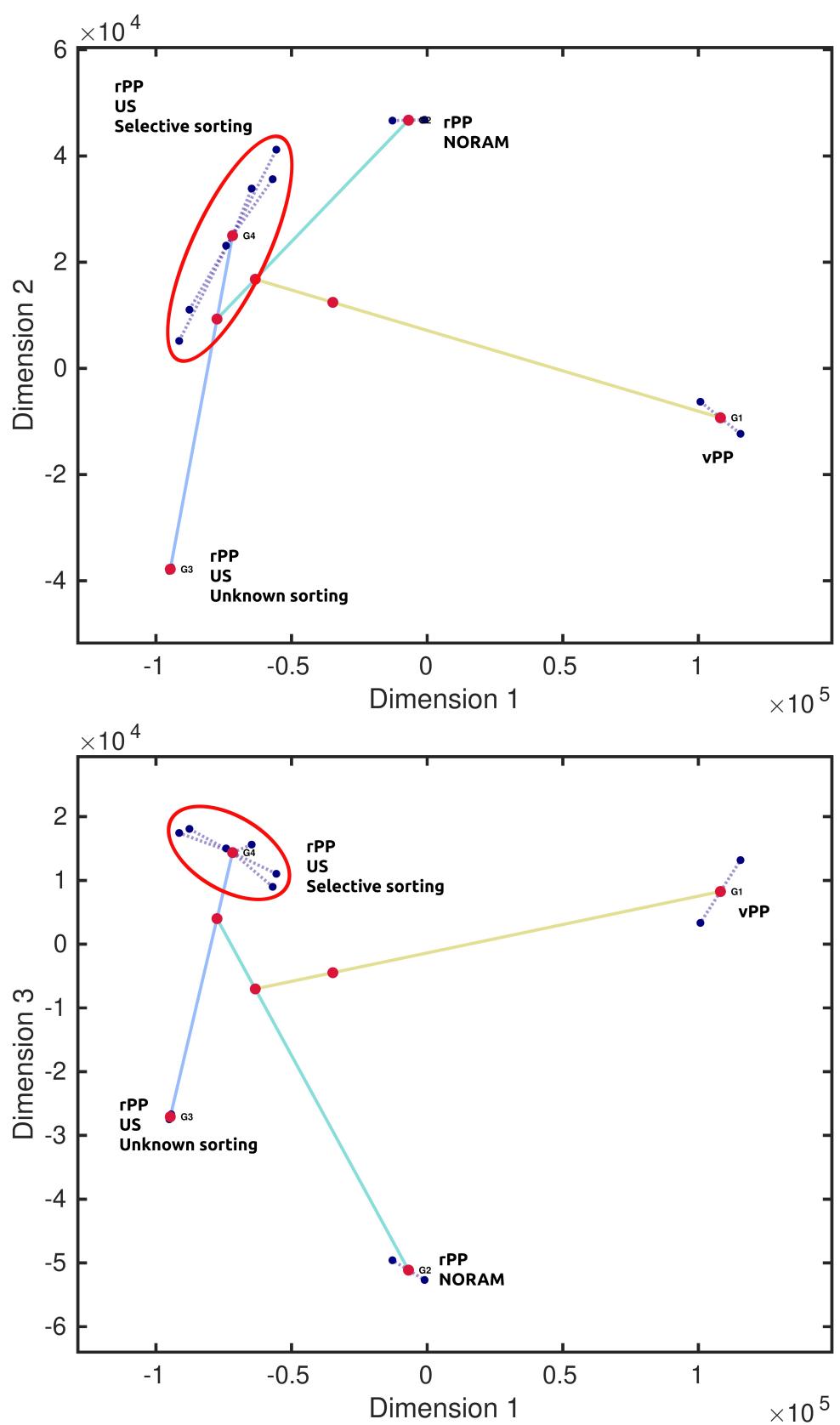


FIGURE 5.32 – Cartes des PP recyclés sur les dimensions 1×2 et 1×3 en distance de Levenshtein.

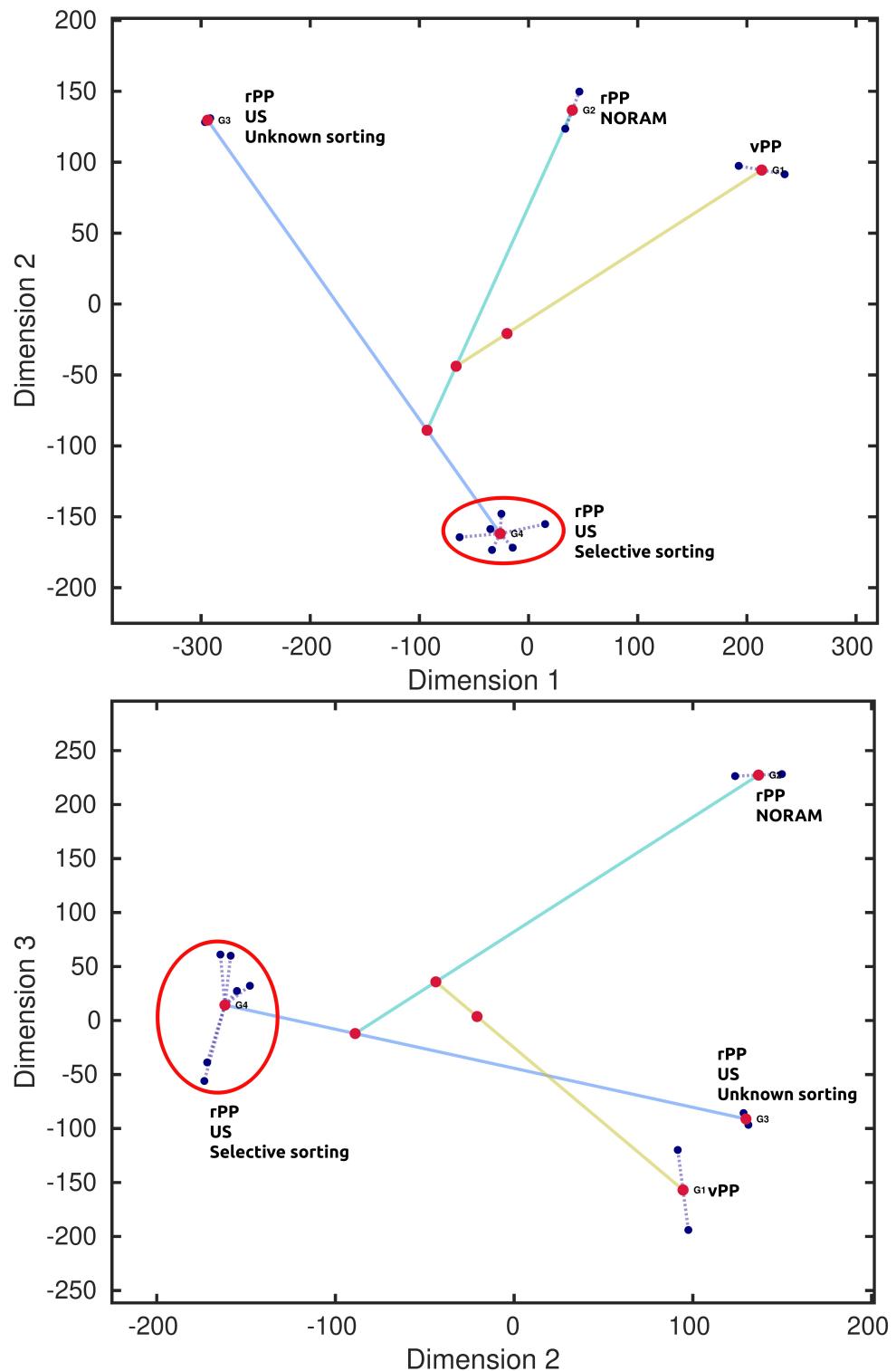


FIGURE 5.33 – Cartes des PP recyclés sur les dimensions 1×2 et 2×3 en distance entropique.

La classification hiérarchique (métrique de Levenshtein) des échantillons est présentée sur la Figure 5.34 sur la base des trois premières dimensions. Les deux critères tendent à donner les mêmes classifications qui préservent ce que l'on sait *a priori* sur les échantillons. Seuls les échantillons centraux (feuilles 2 et 3) sont inversés dans le classement. Le jeu d'échantillons est petit car il est difficile à se procurer, mais il permet de démontrer qu'une méthode mathématique calibrée pour un détecteur de masse donné serait capable de reconstruire une information diffuse relative à des contaminations exogènes liées au tri et l'origine de la matière.

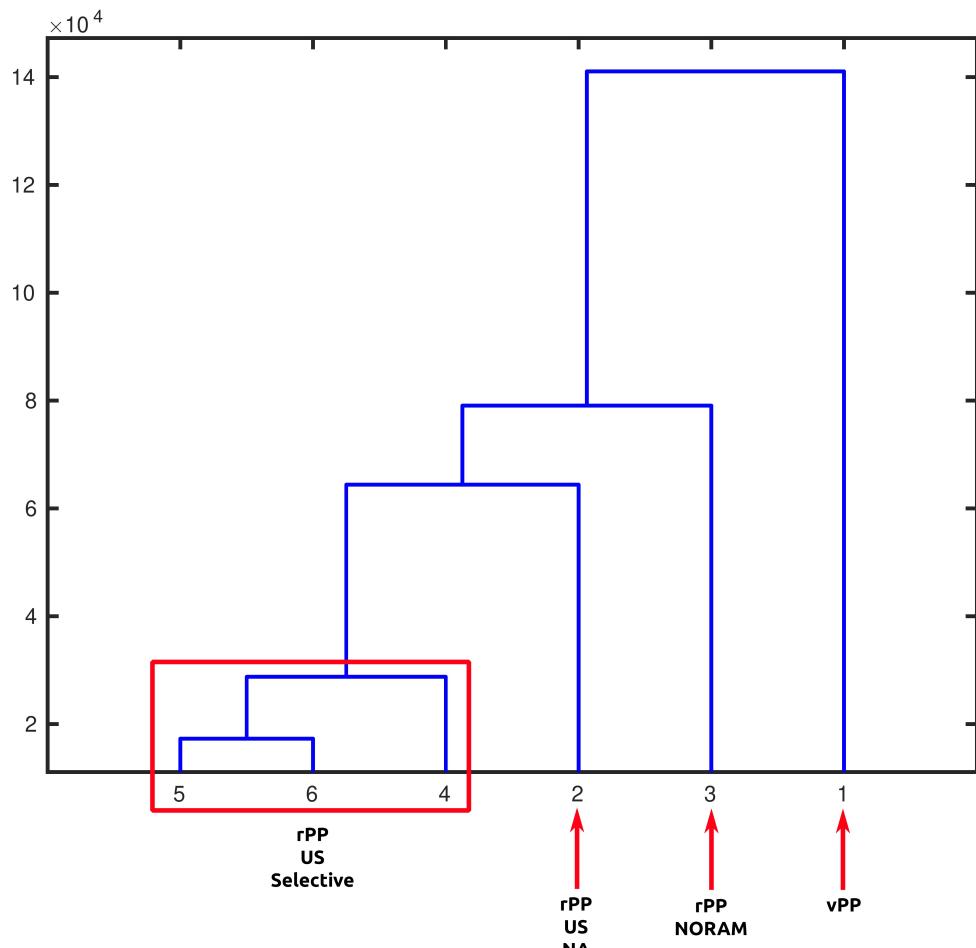


FIGURE 5.34 – Classification *a posteriori* des PP recyclés selon la distance de Levenshtein sur les trois premières dimensions.

Points clefs

Dans l'espace chimique, les critères de collecte et de recyclage entraînent des déplacements chimiques significatifs détectables même sur un tout petit lot d'échantillons. L'espace chimique qui en résulte est bien grand que celui obtenu dans des conditions contrôlées (Partie 5.3.2.2).

5.3.3.4 . Ce qu'il faut retenir de la comparaison des PP et leurs gisements de matières recyclées

A la différence des méthodes de chimiométrie conventionnelles, la méthode de comparaison de chromatogrammes basées sur l'information chimique contenue et non-partagée entre des échantillons permet d'analyser des échantillons de complexité arbitraire sans aucune donnée *a priori* sur la nature des substances présentes (endogènes issues de réactions, exogènes issues de contaminations croisées ou de mélanges). La comparaison ne requiert que deux échantillons et non des dizaines comme dans l'analyse en composantes indépendantes (Partie 2.3.2.2). Le pouvoir de discrimination des méthodes a rapidement dépassé nos attentes initiales, mais à une condition, qui la rapproche d'une certaine manière des méthodes qui demandent un apprentissage ou une projection dans l'espace des données. Les marins avaient l'habitude de reconstruire des cartes grossières sur la base des temps de parcours, mais ces durées étaient associées à de nombreux aléas (vent, marée) qui ne permettait pas faire coïncider précisément les distances perçues avec les distances réelles. Plusieurs voyages étaient nécessaires pour raffiner les distances et les angles. L'analyse en coordonnées principales (PCoA) démontre que les distances entre les mesures sont plus grandes que les distances entre les échantillons parce qu'un bruit résiduel persiste. Elle joue le rôle de filtre en éliminant la moitié du bruit. L'introduction des répétitions permet de définir aisément un seuil de coupure sur les distances pour lesquelles il est possible de dire que deux mesures même non filtrées ne sont pas équivalentes. Elles ne sont pas suffisantes pour reconstruire une métrique ou une carte dans un espace donné. La spatialisation de l'information chimique est particulièrement complexe, car il dépend des détails de la carte que l'on cherche à reproduire. Nous avons montré que des petits effets générateurs de substances pouvaient être détectés et tendent à étendre l'espace le long d'une même dimension. L'introduction de matières recyclées, décontaminées ou non, étend encore plus cet espace et ouvre de nouveaux territoires inconnus.

Les critères de dissimilarités que nous avons construits sur la distance de Levenshtein (Équation 2.6) et d'information mutuelle (Équation 4.4) sont des grandeurs extensives qui augmentent en moyenne (dans un ensemble statistique *ad hoc*) avec la longueur du signal chromatographique et le nombre de différences. Pour un protocole d'acquisition donné, ils mesurent bien un certain nombre de différences ou de mutations qui ne peut que croître :

- quand deux échantillons ou gisements sont mélangés
- quand des substances sont introduites par ajout volontaire ou non (ex. transferts)
- quand des réactions chimiques se produisent.

La signature est construite indépendamment d'une analyse à l'échelle optimale des chromatogrammes (Figure 3.4). Elle préserve ainsi une propriété importante : la matière recyclée décontaminée ressemble plus à de la matière recyclée qu'à de la matière vierge. La trace résiduelle des contaminants si elle est détectable reste encore présente malgré l'élimination de 99% ou 99.9% des substances. Le nombre de dimensions est largement insuffisant pour expliquer tous les effets qui sont très probablement multifactoriels et associés à de très nombreuses substances notamment dans les polyoléfines. La modification du profil de la forêt de pics est toutefois suffisante pour constituer une nouvelle signature. Les deux critères de distance tendent à mesurer les différences en termes de fragments détectés. Les méthodes de comptage bien que différentes mesurent des dissimilarités : nombre d'opérations pour transformer un chromatogramme en un

autre pour le critère de Levenshtein et la variation du nombre de textes possibles si on ajoute ou enlève les fragments à un des termes de la comparaison. L'accumulation d'échantillons permet d'améliorer la reconnaissance d'échantillons semblables ou différents. Tout comme lire plus de livres permet d'aider la compréhension générale de nouveaux textes, l'introduction d'échantillons indépendants aide l'interprétation. Dans son analyse pionnière, [Shannon \(1951\)](#) démontre que tout lecteur a une connaissance statistique suffisante pour prédire la 101^{ème} lettre de toute série constituée de 100 lettres échantillonées aléatoirement de la littérature. Le concept d'information mutuel est probablement le plus général, mais ses performances dépendent de statistiques longues, qui ne sont pas nécessairement disponibles sur des petites séries.

Points clefs

L'accumulation d'échantillons améliore grandement leur différenciation par rapport la comparaison d'une seule paire. Elle permet de remplacer la différence initialement vue comme une distance chimique le long d'une ou plusieurs dimensions qui rapproche ou éloigne les échantillons présentant les mêmes déplacements. Par définition, les échantillons dans la même région de l'espace ont la même signature. Les deux critères retenus (distance de Levenshtein et entropie en excès) génèrent des déplacements chimiques significatifs et exploitables sur des PP recyclés.

5.4 . Application aux polyesters vierges, recyclés et décontaminés

5.4.1 . Vue générale

La partie précédente a montré que la présence de NIAS générés dans des conditions industrielles développaient des différences informationnelles détectables et mesurables. Quand les différences sont faibles un filtrage important est attendu pour extraire et reconstruire la distance associée à un effet isolé. Il est difficile d'établir une correspondance entre distances et nombres de fragments détectés en GC-MS. Une base de calcul approximative de quatre segments par pic (Figure 3.5) et une association de 5 fragments minimum par substance donnent une équivalence de 500 substances pour une distance de 10^4 avec la distance de Levenshtein. Sur cette base, les PP recyclés comptabiliseraient plus de 5000 substances (Figure 5.21). De la même manière, le recyclage du PET est associé à une complexification de la composition des gisements et matériaux comme montré sur la Figure 5.35. Nous pouvons observer que le rPET contient un ensemble supplémentaire de composés formant un massif large dans le centre du chromatogramme ainsi qu'une augmentation significative des concentrations de trois composés en fin d'analyse. Les estimations sont vertigineuses, mais confirment la complexité des analyses bien au-delà des capacités actuelles d'analyse. Le PET est recyclé à grande échelle depuis deux ans pour le contact alimentaire. La demande actuelle est si forte que le PET recyclé (rPET) est bien plus cher que le PET vierge. Les applications pour les eaux minérales et les sodas dominent, mais d'autres applications alimentaires de type barquettes thermoformées prennent de plus en plus d'importance. Il y a une forte concurrence entre les différents usages du PET : produits cosmétiques et détergents. A l'heure actuelle, seules les applications mettant en jeu des produits aqueux sont jugées sûres pour un contact en 100% rPET. Les applications à haute température, boissons alcoolisées et aliments gras requièrent généralement une barrière fonctionnelle ou une dilution avec de la matière vierge. L'orientation des gisements vers des usages sûrs ou acceptables légalement est aujourd'hui indispensable, mais est compliquée par les tensions sur le marché et ses fluctuations.

Un total de 124 échantillons ont été collectés sur les marchés européens, asiatiques et américains (Nord et Sud), 11 échantillons non décontaminés ainsi qu'une référence d'article en matière recyclée ont été prélevés. Tous les échantillons ont été soumis aux mêmes protocoles d'extraction et d'analyse chromatographique que pour les échantillons de PP (extraction, séchage et reprise avec une solution d'étalons internes). Un classement sur les critères de traçabilité est proposé sur la Figure 5.36. Les critères sont organisés selon leur nature :

- état (granulé vierge, recyclé chimiquement, recyclé mécaniquement ou mis en forme)
- origine (Europe, Asie, Amérique du Nord, Amérique latine, États-Unis ou inconnu)
- système de tri (non applicable, retour direct par le consommateur, tri en centre manuel, sélectif ou mixte)
- décontamination (non applicable, chimique, non décontaminé, procédé 1 à 4, polycondensation en phase liquide)

Les références sont des PET standards vierges auxquels tous les échantillons peuvent être comparés. Il faut bien comprendre que les échantillons analysés et leurs distances représentent des différences entre échantillons et non entre gisements. La matière solide recyclée est particulièrement variable avant décontamination. La décontamination, notamment mécanique, contribue

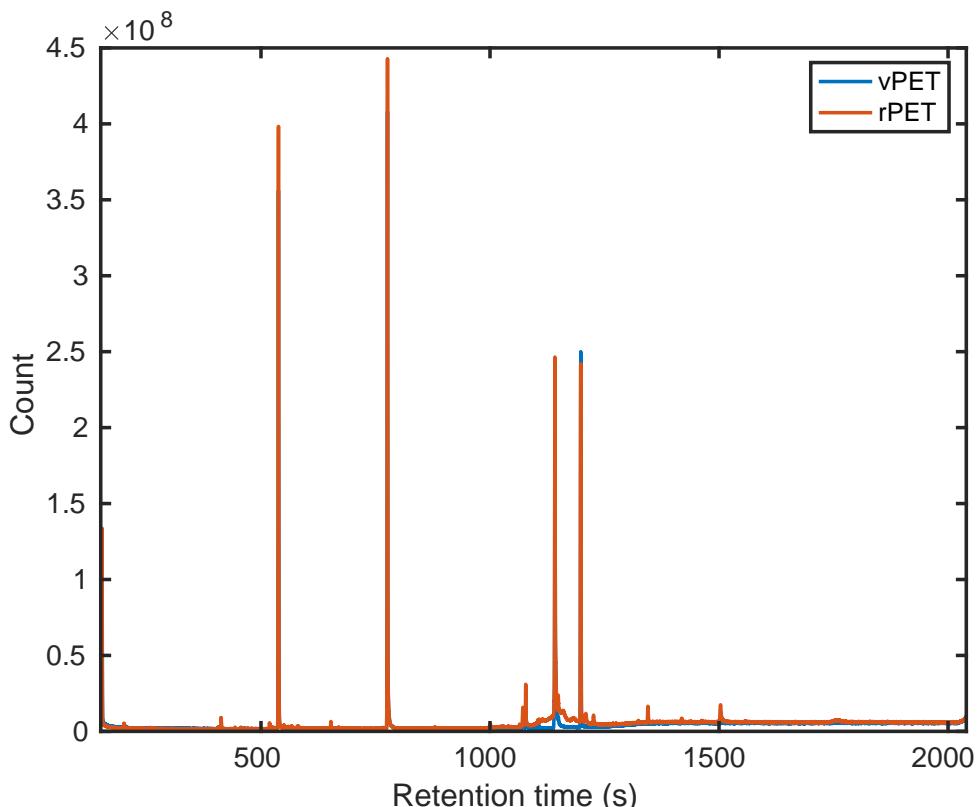


FIGURE 5.35 – Comparaison des TIC issus de l’analyse d’extrait d’un vPET et d’un rPET en GC-MS. Le début du chromatogramme est identique tant en variété des composés qu’en concentration, nous observons l’apparition d’un massif de pics larges signant pour la présence de nombreux composés de concentrations semblables, sur le plateau lié au saignement les pics sont grandement renforcés.

lors des étapes de lavage (généralement à $T_g + 10K$) à éliminer quantitativement les contaminants mais également à les redistribuer. La trituration mécanique puis la granulation contribuent à diluer les contaminants de forte masse moléculaire comme les colorants et les azurants. D’après la littérature, les résidus du rPET sont très nombreux. Outre les résidus de polymérisation, produits d’hydrolyse, de transestérification, nous retrouvons de nombreux composés provenant des différents produits qui ont été à leur contact lors de l’usage (remplissage à chaud, stockage) et du lavage avant recyclage. La dispersion des contaminants des étiquettes, colles et encres est d’autant plus inévitable que les flocons subissent une étape de séchage servant à éliminer l’eau avant stockage et granulation.

La capacité à classer les gisements suivant les critères importants de la filière : reconnaissance de la matière recyclée, de sa décontamination et éventuellement de son origine ont été testés à l’aide d’un échantillonnage aléatoire suivant la règle des quotas. Deux tiers des échantillons sont utilisés pour paramétriser l’espace à partir des distances paires à paires. Le tiers restant est projeté sur cet espace à l’aide des vecteurs propres qui le définissent. La rotation des échantillons d’apprentissage et de validation est répétée un millier de fois pour estimer l’incertitude associée à l’espace topologique qui permet de séparer ces échantillons. Parce que les échantillons sont indépendants, une part de l’incertitude liée aux gisements eux-mêmes est également prise en compte.

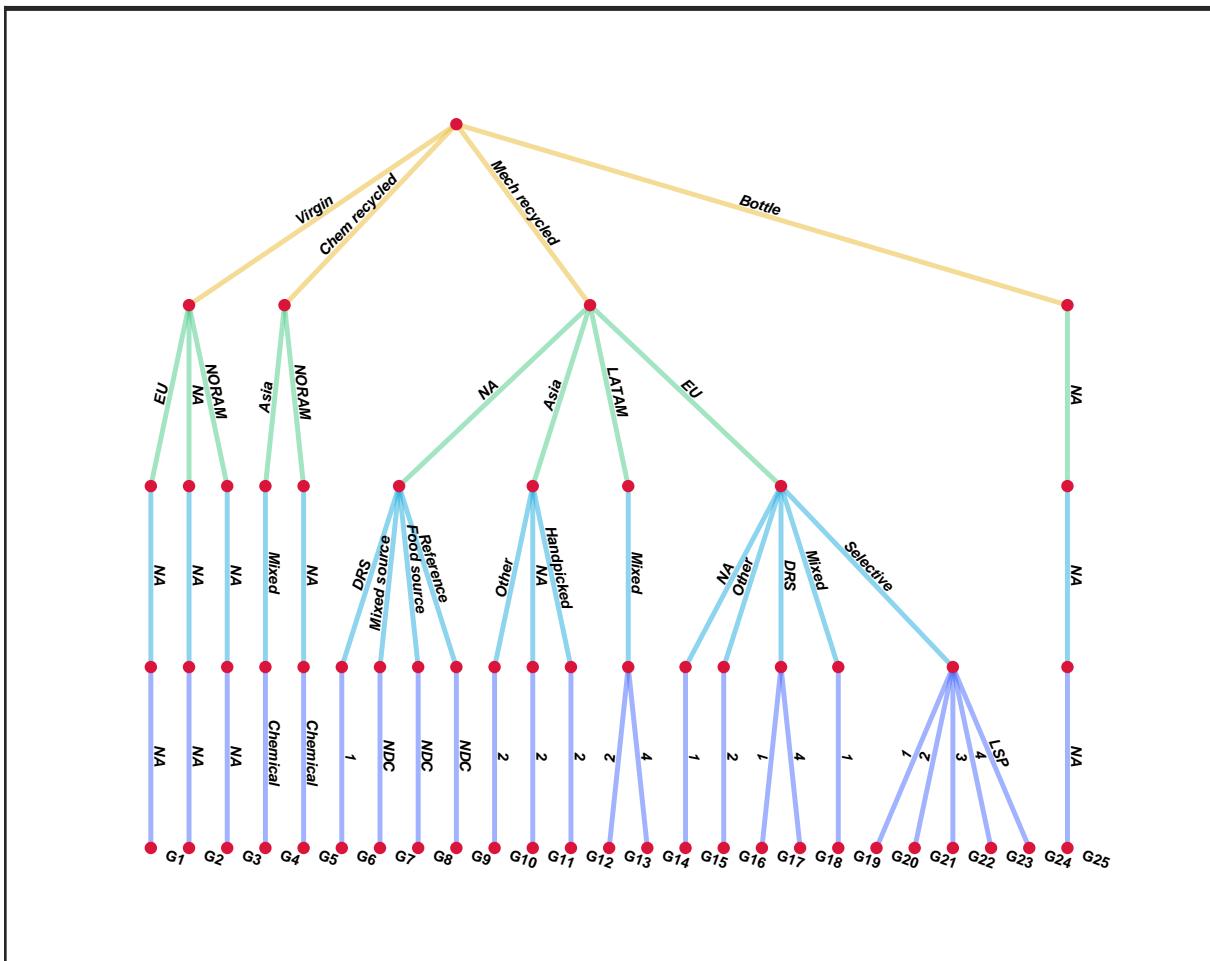


FIGURE 5.36 – Classification *a priori* des PET recyclés selon les critères de traçabilité.

Points clefs

Les chromatogrammes de PET ne sont pas comparables aux PP. Ils sont globalement plus pauvres et la signature est associée à de petites variations des IC relativement proche du critère de rejet du bruit du détecteur.

5.4.2 . Calibration de la cartographie du marché circulaire mondial des PET

Les 124 échantillons recyclés et décontaminés du marché mondial ont été préparés en double pour les analyses, les 11 échantillons non décontaminés et l'article commercial en matière recyclée une seul, totalisant ainsi un jeu de données de 260 chromatogrammes. Ces chromatogrammes ont tous été comparés deux par deux pour recréer une cartographie étendue de l'économie circulaire associée au PET. L'analyse en coordonnées principales fournit alors au maximum un espace de dimension 259 difficilement exploitable et contenant du bruit pour la plupart d'entre elles. La Figure 5.37 donne le carré de la distance à la moyenne (MSD) des dimensions cumulée par rapport à la MSD totale de toutes les dimensions pour les critères de Levenshtein et en entropie. La cassure de la pente s'opère à partir de 16 dimensions pour la distance de Levenshtein, cumulant 75% de la MSD totale. Pour l'entropie la cassure se situe à 4 dimensions pour une MSD de 10% seulement. Toutes les dimensions orthogonales au-delà ne contiennent que du bruit n'expliquant pas les différences entre échantillons.

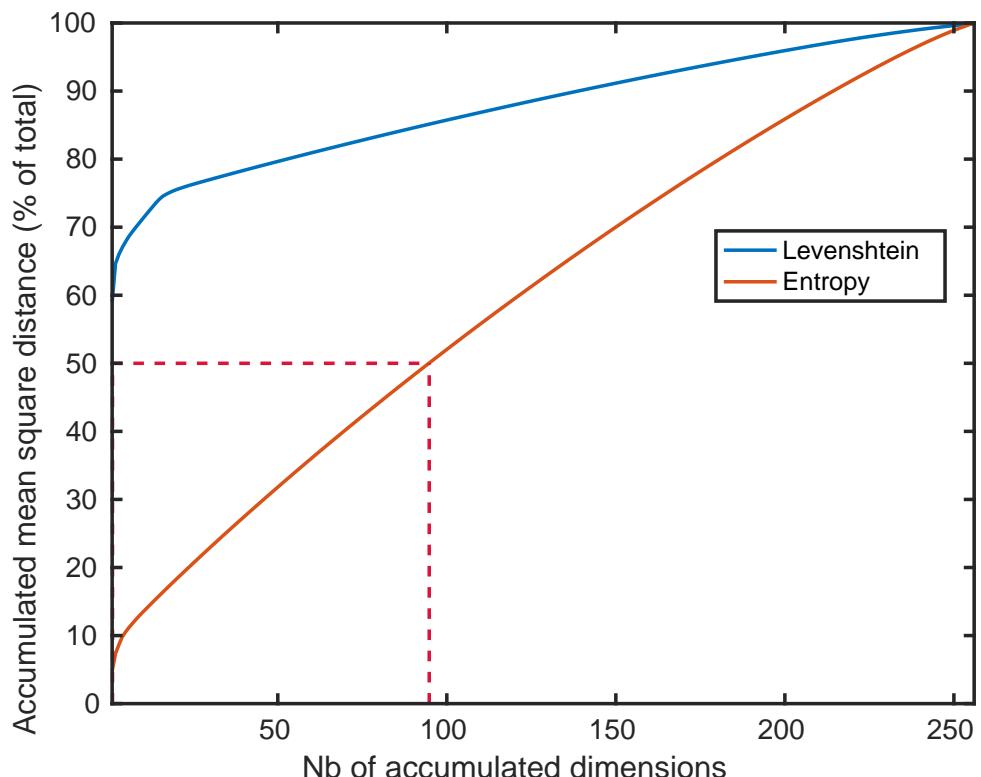


FIGURE 5.37 – MSD du cumul de dimension en pourcentage du MSD total pour les critères de Levenshtein et en entropie appliquée à l'économie circulaire du PET.

Nous avons utilisé les répétitions de préparation des échantillons pour estimer une incertitude (distance minimale) par dimension en deçà de laquelle les échantillons ne sont pas distinguables. Les distances entre échantillons (moyennées entre les répétitions) sont comparées à cette incertitude. La Figure 5.38 recense la proportion de distances par dimension ne dépassant pas cette incertitude et nous définissons une dimension d'intérêt toute dimension dont le premier centile signifiant est inférieur à 0.50. Nous prendrons cependant quelques dimensions supplémentaires jusqu'au centile 0.60, le coût de calcul associé est réduit. Nous retiendrons ainsi 21 dimensions pour la distance de Levenshtein, indiquant que l'information est très diluée avec ce critère. Pour l'entropie, nous nous contenterons de 7 dimensions.

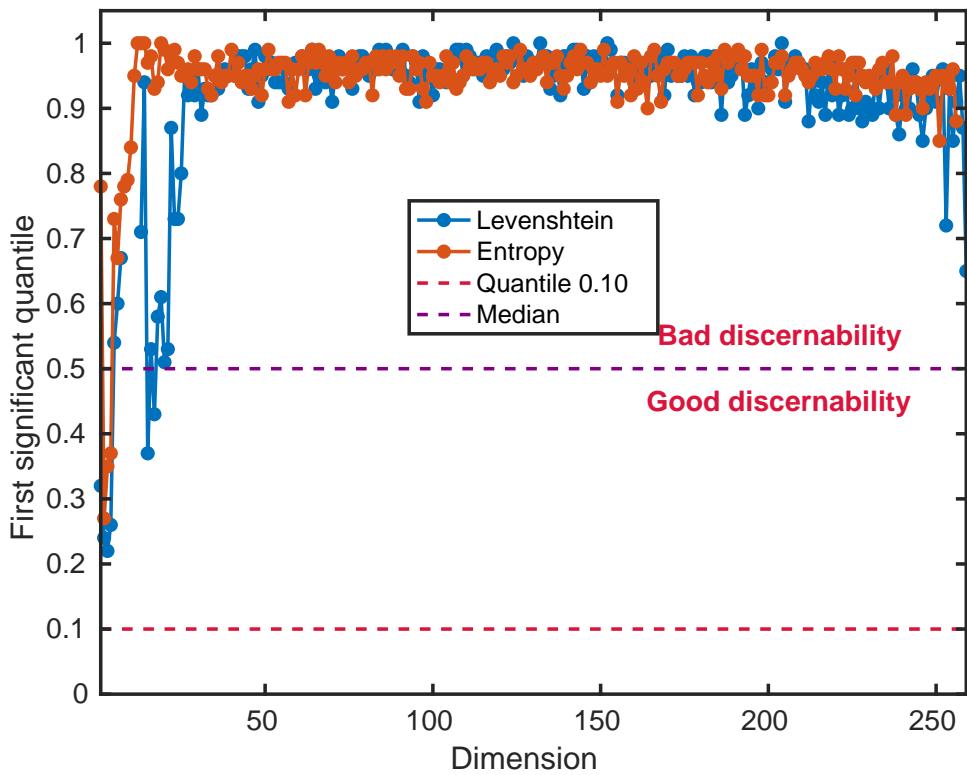


FIGURE 5.38 – Premier quantile significatif par dimension selon les critères de Levenshtein et en entropie pour l'économie circulaire du PET. Toutes les dimensions associées à un centile inférieur à 0.50 sont capables de discriminer efficacement des échantillons.

Points clefs

Le plan d'échantillonnage n'est en soi pas représentatif de la globalité du marché mondial pour le contact alimentaire mais tend à estimer sa variabilité. Les causes possibles de la variabilité des gisements ne sont pas toutes connues, cependant les effets liés au mode de tri, à l'origine géographique et au type de décontamination ont pu être contrôlées. Suivant ce classement, 136 échantillons ont été répartis dans 25 groupes (G1-25) et sur lesquels ont été réalisées un total de 260 mesures.

5.4.3 . Reconstruction de la traçabilité et cartographie du marché circulaire mondial des PET

Nous rappelons que nous avons à disposition quatre éléments de traçabilité sur les échantillons de PET prélevés sur le marché mondial. Nous avons appliqué une analyse de la variance présentée en Figure 5.39 sur ces quatre effets isolés (4-way) et imbriqués (nested) sur les critères de Levenshtein et en entropie. Nous constatons que le tri S a le plus de probabilité de s'exprimer dans les dimensions 5 et 6, la décontamination D dans la dimension 6 pour le critère de Levenshtein, les autres éléments de traçabilité étant minoritaires. Pour le critère entropique, S s'exprime dans la dimension 6 et 7, D dans la dimension 4.

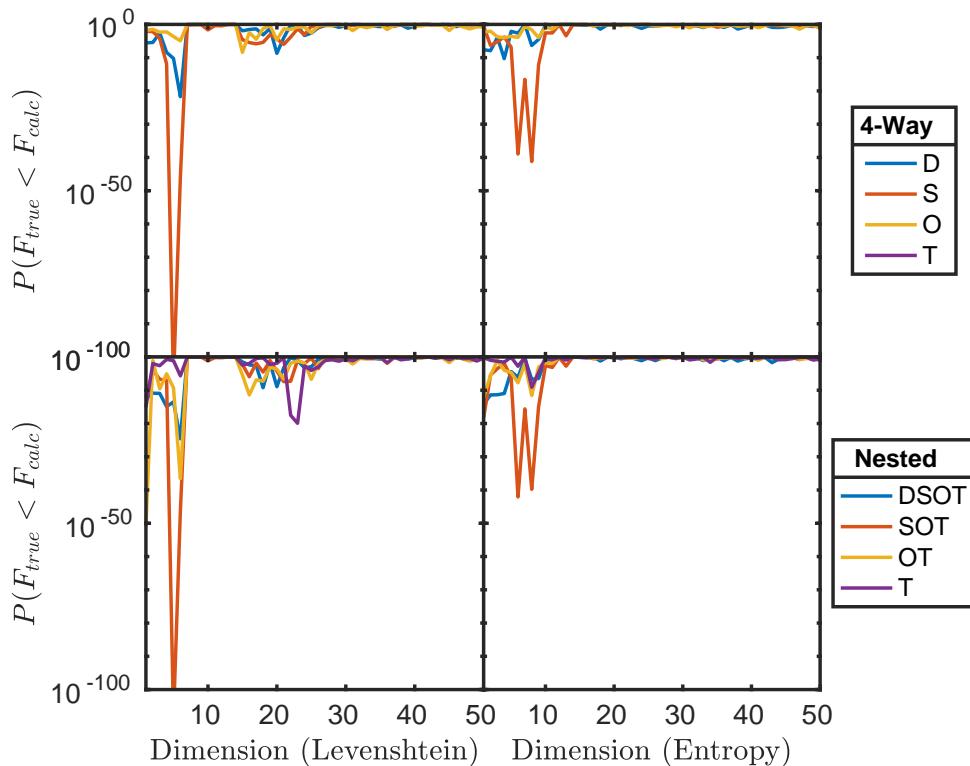


FIGURE 5.39 – Premier quantile significatif par dimension selon les critères de Levenshtein et en entropie pour l'économie circulaire du PET. Toutes les dimensions associées à un centile inférieur à 0.50 sont capables de discriminer efficacement des échantillons.

Les Figures 5.40 et 5.41 montrent les projections sur les deux premières dimensions selon les critères de Levenshtein et en entropie, la seule séparation qui s'opère est celle entre les échantillons décontaminés et non décontaminés. Les projections sur les dimensions extraites par l'ANOVA n'apportent pas plus de séparation. Les échantillons appartenant au même groupe de traçabilité (liens en pointillés sur les cartes) sont épargnés et leur barycentre n'a pas de réalité physique, cela confirme que l'ANOVA basée sur les groupes *a priori* n'a pas de sens sur ce jeu d'échantillons.

La variabilité des échantillons, même ceux possédant les mêmes critères de traçabilité, ne permet pas de redonner du sens aux dimensions. La composition d'un échantillon n'est pas représentative d'un gisement dans son ensemble et encore moins de son évolution dans le temps en fonction des arrivages de produits à recycler. Par la suite nous essaierons uniquement de classer les rPET par rapport aux références stables que sont les vPET et potentiellement les rPET issus d'un recyclage chimique, la classification selon l'origine et le procédé de tri étant trop aléatoire. La Figure 5.42 présente la distance les échantillons de rPET aux références vPET avec le critère entropique dans un diagramme de Pareto, plus la distance est courte plus l'échan-

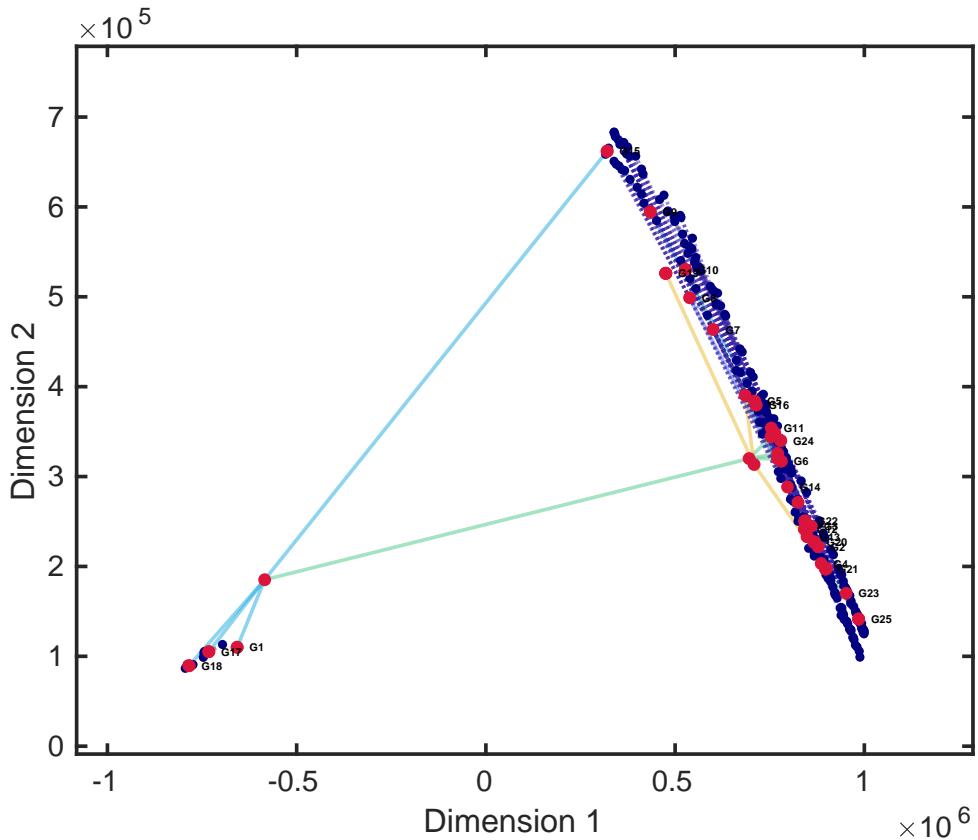


FIGURE 5.40 – *Projection des PET sur les dimensions 1 × 2 selon le critère de Levenshtein. Les échantillons non décontaminés sont séparés du reste et se trouvent dans la partie gauche de la carte.*

tillon contient une information proche d'un matériau vierge. Nous pouvons observer que les échantillons de rPET issus d'un recyclage chimique ne sont pas nécessairement plus proches des vPET contrairement à ce que nous pouvons attendre de ce procédé coûteux. Les échantillons non décontaminés se trouvent à une grande distance des vierges et sont accompagnés d'échantillons recyclés mécaniquement et décontaminés qui seraient évalués comme impropre au contact alimentaire.

La Figure 5.43 représente une classification *a posteriori* des échantillons de PET dans l'espace des sept premières dimensions du critère entropique, les rPET non décontaminés sont bien isolés du reste du jeu d'échantillons. Les échantillons de rPET associés à des vPET ou des rPET recyclés chimiquement ont des chances d'être compatibles avec une utilisation alimentaire.

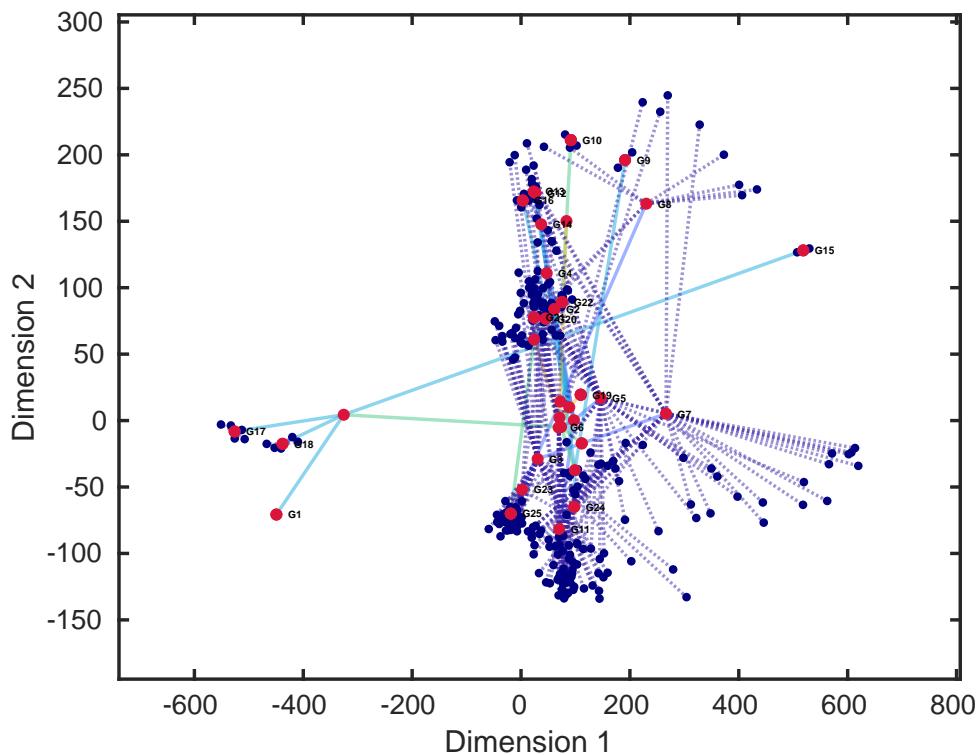


FIGURE 5.41 – Projection des PET sur les dimensions 1 × 2 selon le critère entropique. Les échantillons non décontaminés sont séparés du reste et se trouvent dans la partie gauche de la carte.

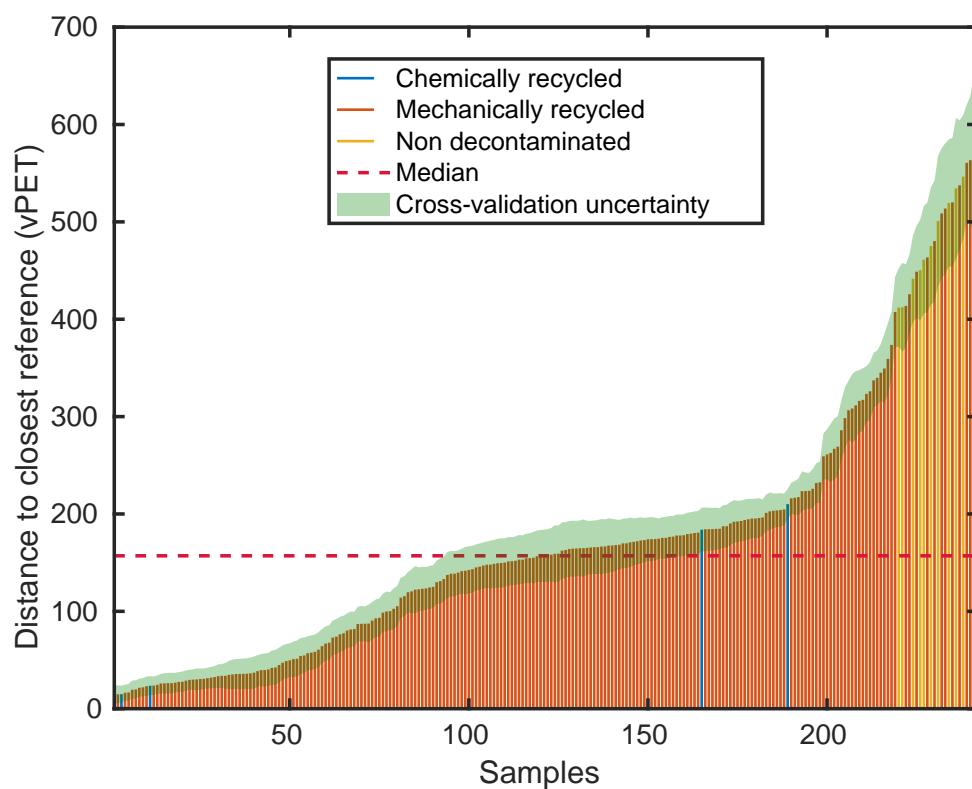


FIGURE 5.42 – Classification a posteriori des PET recyclés selon la distance de Levenshtein sur les trois premières dimensions.

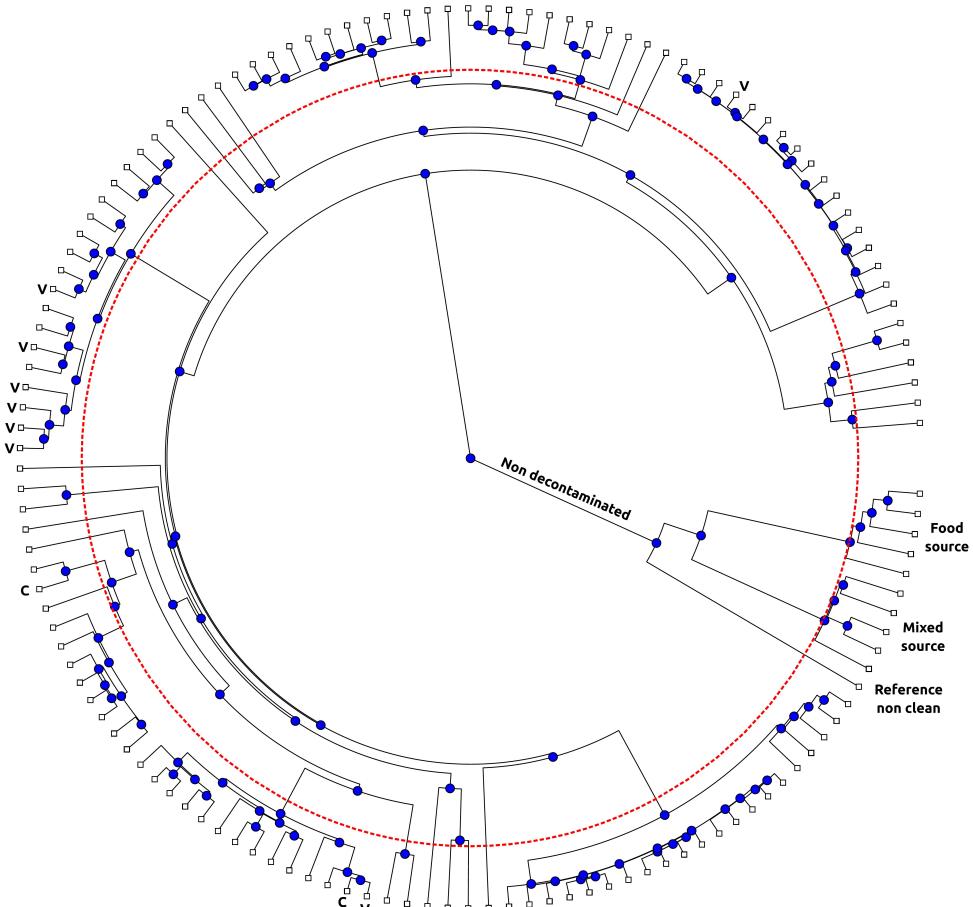


FIGURE 5.43 – Classification *a posteriori* des PET vierges et recyclés selon la distance entropique sur les sept premières dimensions. Les échantillons vierges (V), recyclés chimiquement (C) et nondé-contaminés sont indiqués en bout de branche. Le cercle en pointillés indique la distance de référence depuis le bord du disque dans cet espace correspondant à l'incertitude entre deux répétitions et égale à 50. La distance du centre au bord représente la distance maximale entre tous les échantillons et vaut 400.

Points clefs

L'espace chimique développé a été interprété dans la logique de reconstruction de la traçabilité. Les distances entre échantillons sont très grandes et une carte près de 100 fois plus grande que la carte obtenue pour les PP (Partie 5.3.2.3) a été obtenue. Elle montre une très grande diversité de déplacements dans beaucoup de dimensions qui laissent penser que de nombreux effets non traçables (dans le cadre de l'étude) comme le type de produit précédemment en contact (eau, sodas, jus d'orange, détergents, shampoings) ou le nombre de recyclages affectent la signature chimique détectée. Parce qu'il s'agit dans beaucoup de cas de grades proposés pour le contact alimentaire, les deux critères montrent que les PET recyclés pour le contact alimentaire ne sont pas identiques aux PET vierges. L'analyse stochastique montre que tous les grades ne sont pas équivalents et qu'un critère de pureté peut être reconstruit a priori. La méthode est capable de mettre en évidence des défauts de décontamination et montre paradoxalement que les échantillons recyclés chimiquement ne sont pas les plus propres.

5.4.4 . Ce qu'il faut retenir de la comparaison des gisements de PET recyclés et de leur décontamination

5.4.4.1 . Ce qui change quand la cartographie chimique est étendue à toute une filière

Les résultats sur un vaste échantillon de PET issus de déchets ménagers et ou d'une filière de recyclage sur le marché européen sont imprégnés d'une diversité d'information chimique bien supérieure à celle de jeux d'échantillons plus restreints sélectionnés ponctuellement (Partie 5.3.3 sur les rPP). Nous aurions pu penser que la complexité des différences est proportionnelle à la complexité initiale des extraits de polymères. Le PP offre intrinsèquement la plus grande complexité avec de nombreuses possibilités de créer de nouveaux composés et pics : la scission d'un oligomère donne en cascade deux autres oligomères avec des niveaux possibles d'oxydation variés. Cet effet est décelable, mais il est inférieur aux différences observables dans une économie circulaire. Il y a un ordre de grandeur en termes de différences entre les premières dimensions des cartes des rPP et des rPET selon le critère de Levensthein (Figures 5.32 et 5.40). Les estimations donnent des signatures qui reposent sur au moins une centaine de substances. Ce nombre est difficile à établir car toutes les mesures ne sont pas réalisées au même moment et la dérive de la colonne chromatographique pollue les comparaisons par des composés issus de la colonne. Les critères appliqués aux signaux chromatographiques (Équation 5.9) permettent de rejeter statistiquement 95% des signaux parasites, mais nous ne pouvons pas écarter que des faux positifs soient comptés parmi les différences. Les répétitions et le filtrage permettent de réduire dans une certaine mesure l'incertitude, la distance quadratique est ainsi réduite de 50% à 70% entre échantillons (Figures 5.29 et 5.37). Parce que les programmes de température et le type de colonne sont similaires entre les échantillons du PET et du PP, les biais intrinsèques sont incapables d'expliquer l'augmentation considérable de l'étendue des dissimilarités des PET issues de l'économie circulaire.

Points clefs

La filière de recyclage du PET est particulièrement complexe. Elle est d'abord appliquée à l'échelle mondiale avec un taux d'incorporation variable de matériaux en contact avec des produits non alimentaires. Les procédés de collecte, tri et recyclage restent peu standardisés et peu caractérisés. Sans hypothèse supplémentaire, les critères proposés montrent un espace chimique très vaste.

5.4.4.2 . Espace chimique et flux d'information dans une boucle de recyclage

L'espace chimique H (ou bruit de fond) qui contrôle les contaminations fortuites dans l'économie circulaire n'est pas complètement caractérisé, mais il couvre potentiellement la totalité des substances utilisées dans les produits et les matériaux, qu'ils soient utilisés pour un contact alimentaire ou non (ex. encres d'impression). Cette représentation informationnelle est indépendante des quantités présentes dans les boucles de recyclage ou dans leur voisinage. Ce qui est inconnu, c'est le "flux d'information de la circularité" en provenance de H et qui affecte la contamination d'un gisement/matiériaux lors de son recyclage ou de son réemploi. Ce concept peut être mathématisé et être défini comme le flux d'un champ au travers d'une courbe. Cette quantité est un scalaire noté dans ces travaux par la lettre hébraïque \daleth (daleth). Parce que \daleth n'est pas défini par rapport à sa circulation, il ne comptabilise pas les contaminants transportés par l'échantillon (son flux est parallèle à la trajectoire), seule la contamination exogène (orthogonale) est considérée. Les effets exogènes sont nombreux (contaminations croisées entre matériaux, produits et fluides de lavage). Si deux effets, appelés e_1 et e_2 , sont considérés on obtient le flux d'information issu du bruit de fond et qui traverse de toute part l'itinéraire technologique, vu comme la courbe Γ , est donné par :

$$\nabla = \iint_{\Gamma} (P de_2 - Q de_1) \quad (5.15)$$

où P et Q représentent une mesure du nombres de substances dans l'espace (e_1, e_2) de \mathbb{R}^2 . Dans le cadre de mélange et de recyclages multiples, les flux sont additifs $\nabla = \sum_i \nabla_i$ et donc extensifs avec la longueur de Γ . L'inégalité triangulaire permet d'estimer par défaut ∇ à partir des distances paires à paires filtrées, δ . Par exemple le recyclage d'un vPET en rPET peut est illustré par la Figure 5.44 :

$$\mathfrak{d} \geq \tilde{\delta}(rPET, vPET) \approx Card(rPET) - Card(vPET) \quad (5.16)$$

où Card représente une quantité associée à une entité donnée, dans notre cas l'information contenue par le vPET ou le rPET et dont la différence traduit l'ajout ou le retrait d'information au cours du chemin de recyclage.

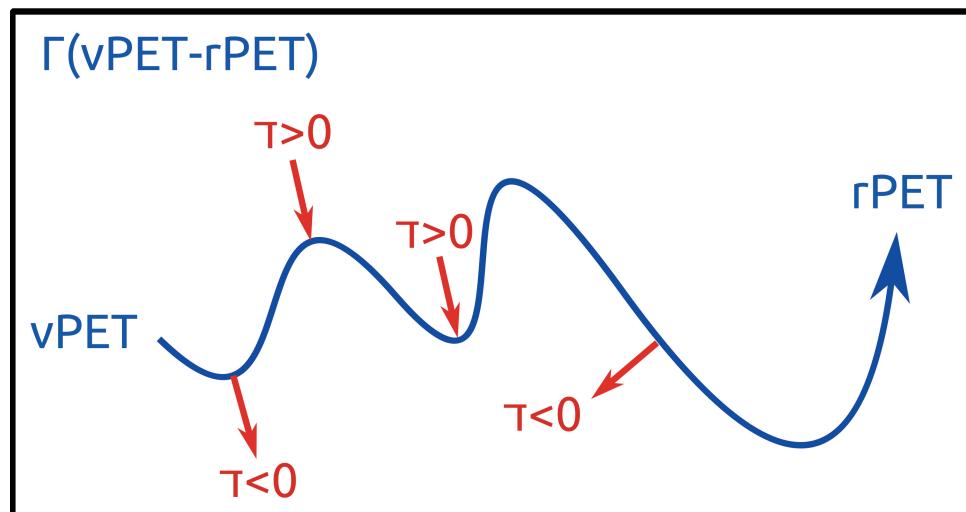


FIGURE 5.44 – Exemple de chemin Γ de recyclage de PET avec plusieurs ajouts et retraits d'informations \top perpendiculaires au trajet.

Les résultats montrent qu'il existe une hiérarchisation des effets. En prenant les distances paires à paires, la hiérarchisation suivante est intuitive $\mathsf{chimique} < \mathsf{mécanique} < \mathsf{mélange} \ll$

$\Gamma_{contaminé}$. Il n'est pas possible de préciser plus les effets faute d'une traçabilité étroite des échantillons. La validation croisée appliquée avec un échantillonnage sur la méthode des quotas montre que cette hiérarchisation est indépendante des échantillons utilisés pour filtrer les flux. La valeur de Γ et ce quel que soit le critère retenu et avec un petit nombre de dimensions (inférieur à 7) permet d'identifier à l'aveugle l'origine d'un PET sur la base d'une comparaison simple avec un PET de référence. Pour le PET, la méthode permettrait de faire une identification en quelques heures (le facteur limitant est le temps d'extraction) tout comme la méthode RT-PCR (reverse transcription-polymerase chain reaction) en temps réel qui permet de détecter la COVID-19. L'analogie est confirmée par le texte de l'agence de l'ONU AEIA qui rappelle que la technique de détection des virus pathogènes (Ebola, Zika, MERS, SRAS, COVID-19) repose sur une technique d'identification de séquences spécifiques (équivalents aux contaminants de l'économie circulaire) dans une forêt de codons développée initialement par l'industrie nucléaire. "À mesure que les copies des séquences de l'ADN viral sont produites, les marqueurs se fixent sur les brins d'ADN et émettent une fluorescence qui est mesurée par l'ordinateur de l'appareil. Les résultats s'affichent en temps réel à l'écran. L'ordinateur effectue un suivi de la quantité de fluorescence dans l'échantillon à la fin de chaque cycle. Lorsque le niveau de fluorescence dépasse un certain seuil, la présence du virus est confirmée. Le nombre de cycles nécessaires pour atteindre ce seuil permet également aux scientifiques d'estimer la gravité de l'infection : plus le seuil est atteint rapidement, plus l'infection est grave" (Jawerth, 2020).

L'analogie avec la biologique n'est pas abusive ici car les méthodes que nous utilisons pour mesurer les distances entre chromatogrammes sont adaptées des méthodes d'alignement des génomes ou protéomes et mettent en jeu un code. La capacité à retrouver l'origine d'un PET et notamment à identifier une étape de décontamination peut paraître très surprenante puisque que cela à reviendrait à dire que la décontamination efface une partie de l'information alors que les concentrations sont seulement abaissées. Il n'y a pas de paradoxe. On retrouve exactement avec le même problème avec les virus. Ils peuvent ne pas être détectés alors qu'ils sont toujours présents dans le génome, mais ne s'expriment pas (dormance). Le recyclage mécanique n'efface pas l'information, mais peut réduire sa détectabilité et quantification. Si elle met en jeu un mélange avec d'autres gisements, l'information peut également augmenter. Les critères retenus mesurent le bilan entre ces deux effets antagonistes. Le recyclage chimique permet de retirer les contaminants exogènes en phase vapeur avec ou sans craquage, mais peut conduire à d'autres contaminants endogènes. De manière plus générale, le vieillissement des polyesters compensé par des réactions de polymérisation à l'état solide contribue à complexifier l'empreinte chimique. L'ensemble de ces facteurs est mesuré et rapporté à une seule distance. La décomposition des effets suivant les directions causales $e_1, e_2 \dots$ est possible sur la base d'une analyse de la variance imbriquée du type de celle présentée pour le PP dans la Partie 5.3.2.4. Elle requiert toutefois de maîtriser les matériaux entre deux niveaux de sorte que les conséquences cumulées puissent être étudiées sous la forme d'une trajectoire $e_1 \rightarrow e_2 \rightarrow \dots$ (il s'agit d'un exemple de trajectoire Γ). Cette analyse n'a pas été possible avec les échantillons du PET du fait de l'absence de coopération des acteurs de la filière de recyclage.

Tout comme en génétique des populations, l'économie circulaire implique un brassage des gisements. Ils sont amenés à être réutilisés plusieurs fois pour des applications différentes et dans des régions géographiques très éloignées. La mesure de la divergence à un état donné peut fournir une indication du nombre moyen \tilde{n} de passage du gisement dans la boucle de recyclage : incluant contamination, mélange, décontamination, dégradation, etc. Il s'agit ainsi d'évaluer combien de fois \tilde{n} un chemin Γ a été parcouru en supposant une séquence aléatoire mais stationnaire. L'extensivité de Γ suggère que l'âge moyen \tilde{n} est donné par :

$$\tilde{n} \approx \frac{\tau_n}{\tau} \quad (5.17)$$

Cependant les échantillons que nous avons à disposition ne nous permettent pas de confirmer cette réflexion. Cela nécessiterait d'avoir un vPET et au moins son équivalent recyclé pour définir une distance de référence pour un premier cycle pouvant être transposer aux recyclages successifs.

Points clefs

La boucle de recyclage tout comme l'itinéraire technologique peut être vu comme un espace abstrait dans lequel de l'information est échangée avec la matière collectée, triée, lavée, décontaminée. Les étapes ou les cycles successifs représentent autant de chemins parcourus. Les flux sont additifs. Les distances correspondantes entre échantillons tendent donc à être croissantes avec la longueur des chemins. Les nombres sans dimension proposés τ sont à l'origine des déplacements chimiques. La décomposition orthogonale des τ_i (dans l'espace des échantillons) correspond à une factorisation des effets du fait de la pratique industrielle. L'analyse actuelle ne permet pas de retrouver les causes mais permet de détecter le nombre d'effets non corrélés. Nous en avons identifié entre 20 à 30.

5.5 . Discussion Générale

5.5.1 . Préambule

5.5.1.1 . Des valeurs de γ qui sont largement au-delà de nos estimations initiales : des risques émergents probables au-delà de la problématique initiale des NIAS

Les résultats présentés ont deux contributions. La première contribution est méthodologique et permet de traiter à très haut débit des problématiques de NIAS à partir d'extraits de polymères. La méthode est non supervisée et ne requiert en rien l'identification manuelle des NIAS. Les NIAS sont définis par une différence d'empreintes dans un espace transformé du chromatogramme. Le codage en lettres permet de compresser le signal et d'utiliser de très nombreux outils génériques issus du domaine de la biologie ou de l'analyse textuelle pour rechercher une substance spécifique, un mélange de substances, aligner et comparer des extraits. La seconde contribution est liée à l'application de la méthode à l'économie circulaire et à un concept tout à fait original que nous avons introduit par analogie avec les autres activités de l'UMT SAFEMAT 22.07. Les matériaux s'échangent et s'imprègnent de substances à tous les stades de leur cycle de vie. Le recyclage offre des gains d'énergie et de matière significatifs pour les matériaux d'emballage mais au prix d'un risque de contamination pour les aliments au contact des matières recyclées. Au démarrage de la thèse, la problématique se limitait à quelques substances, notamment celles qui pourraient être générées par un manque de contrôle des procédés de fabrication de ces matériaux. En remarquant que l'économie circulaire des matériaux pouvait également s'interpréter en termes thermodynamiques notamment en lien avec la forte expertise de l'UMT en transferts de matière, les travaux ont introduits une mesure globale de l'échange ou de l'apparition de substances sous la forme d'une mesure de l'entropie dérivée de celles utilisées pour décrire les systèmes ouverts. On serait en droit de douter que la combinaison des deux contributions, insuffisamment testées et développées dans le cadre d'une seule thèse, puisse nous apprendre quelque chose sur l'état sanitaire associé à nos boucles de circularité utilisées actuellement sur les marchés européen, américain et asiatique. Il n'y a toutefois aujourd'hui ni d'approche concurrente, ni de suivi publié qui justifieraient que nous fassions l'économie de l'analyse des constats réalisés à partir des nouvelles approches. Le premier constat est que la complexité des compositions des matériaux issus des économies circulaires actuelles est bien au-delà des prévisions initiales et des valeurs rapportées dans la littérature. Les gisements de PET présentent des estimations de cent fois plus grandes que celles rencontrées dans les issus d'une production contrôlée de PP et qui devaient servir de base à la construction d'une stratégie d'identification des NIAS pour le laboratoire national de référence. L'économie circulaire requiert d'identifier et de caractériser beaucoup plus de composés à partir d'un petit nombre d'échantillons. Identifier et quantifier 100 NIAS dans un échantillon est au-delà des capacités des études publiées notamment en chimiométrie (ex. dans l'étude de 12 composés étaient au plus identifiés). Ce n'est pas leur nombre qui présente la difficulté la plus importante, mais leur très grande diversité entre échantillons. Il est peu probable que l'accumulation de chromatogrammes produits dans des conditions contrôlées (alignées, filtrées) permettent l'extraction de sous signaux d'un si grand nombre de NIAS potentiels (des centaines voire des milliers pour une dizaine d'échantillons indépendants). Parce que les matériaux recyclés sont sur le marché et leur utilisation est encouragée voir rendue obligatoire par les nouvelles lois environnementales, il y a une urgence à préciser sur ces NIAS mesurés ici par agrégation peuvent présenter un danger pour le consommateur. Il ne s'agit en effet plus de gérer une contamination fortuite par une substance, mais une contamination systématique des matériaux par des dizaines de substances, dont la composition exacte du cocktail est aléatoire. Nous retrouvons les mêmes préoccupations dans l'analyse des PET post-consumer réalisées par l'institut Fraunhofer ([Franz and Welle, 2020](#)).

5.5.1.2 . ↗ : un concept possible pour gérer l'urgence réglementaire

Le manque de recul des autorités et des laboratoires de référence sur les risques d'exposition cumulée crée l'urgence. Le règlement (EU) 2022/1616 qui est entré en vigueur le 10 octobre 2022 ouvre de manière transitoire à l'exploration de nouvelles technologiques de recyclage et de décontamination ainsi qu'à l'utilisation de matières recyclées autres que le PET.

- L'article 6 impose une traçabilité et un contrôle à la collecte : "Les déchets plastiques sont contrôlés tout au long de la collecte et de la prétransformation au moyen de systèmes d'assurance de la qualité. Les systèmes d'assurance de la qualité : a) garantissent le respect des conditions et des exigences énoncées aux paragraphes 1 et 2; b) garantissent la traçabilité de chaque lot jusqu'au premier tri des déchets plastiques collectés; et c) sont certifiés par un tiers indépendant."
- L'article 7 impose un enregistrement des performances des décontaminations : "Un répertoire des données enregistrées, utilisé pour consigner les informations sur la qualité des différents lots, telles que définies à la section 4.1 de la fiche récapitulative de contrôle de la conformité visée au paragraphe 3, point c), est tenu à jour. Les données conservées dans ce répertoire le sont pendant une période d'au moins cinq ans."
- L'article 8 impose de nombreuses obligations de transfert d'information de la part des transformateurs utilisant de la matière recyclée.
- L'article 9 exige à tout exploitant du secteur alimentaire une traçabilité sur l'usage et qu' "ils ne sont pas contaminés par des matières ou substances autres que celles autorisées par le système de recyclage."

Ces mesures exceptionnelles par leur ampleur doivent être mises en place sans attendre. Le concept ou son interprétation thermodynamique en termes de transfert peuvent y répondre en partie ou totalité.

5.5.1.3 . Éléments de la discussion

L'appropriation par des opérateurs et à grande échelle de nouveaux concepts demandera plusieurs étapes de validation préalables qui seront discutées dans cette partie. Ils sont présentés sous la forme de défis.

Défi 1 : transferts et information mutuels

La capacité de mesurer un transfert de substances à l'aveugle doit pouvoir être montré dans un contexte plus large et mutuel. Deux matériaux qui seraient distincts dans l'espace chimique doivent être vus plus proches s'ils se sont échangés de la matière.

Une étude préliminaire de la contamination entre quatre aliments et leur emballage en contact à un temps de référence puis après un vieillissement accéléré de 20 jours à 60°C. Le Tableau 5.10 résume les aliments ciblés ainsi que les différentes parties de leur emballage associé, l'étude totalise 22 échantillons (11 références et 11 vieillis) qui ont tous été préparés par extraction suivie d'une reconcentration puis d'une analyse en GC-MS. Les chromatogrammes obtenus ont été comparés à l'aide de la méthodologie développée dans ce manuscrit sur le critère de Levenshtein.

TABLEAU 5.10 – *Échantillons utilisés pour l'étude préliminaire de la contamination entre aliments et emballage. Chaque aliment est associé à un ou plusieurs matériaux constituant leur emballage.*

Aliment	Constituants de l'emballage
Céréale	Sachet (PE*), carton
Farine	Papier
Pâtes	Carton, colle, fenêtre (PE*)
Riz	Carton

*PE : polyéthylène

La Figure 5.45 présente la cartographie sur les dimensions 1×2 obtenue après la comparaison des chromatogrammes avec le critère de Levenshtein. Nous pouvons observer le déplacement des chromatogrammes des échantillons avant et après vieillissement. Ces déplacements peuvent signifier à la fois pour la formation et la dégradation de composés, mais aussi pour la contamination mutuelle entre l'aliment et l'emballage. Tous les déplacements ne sont pas équivalents, les plus importants concernent :

- le sachet de céréales, en contact direct avec un aliment gras
- le papier de la farine et le carton du riz qui sont des aliments avec une grande surface spécifique favorisant les échanges de composés.

Cette étude est menée à être étendue à plus d'aliments et emballages avec l'acquisition de points plus réguliers dans le temps. Et les récents développements de la méthodologie de comparaison pourraient permettre la mise en évidence l'appariement entre certaines dimensions et des effets notoires du vieillissement accéléré des emballages.

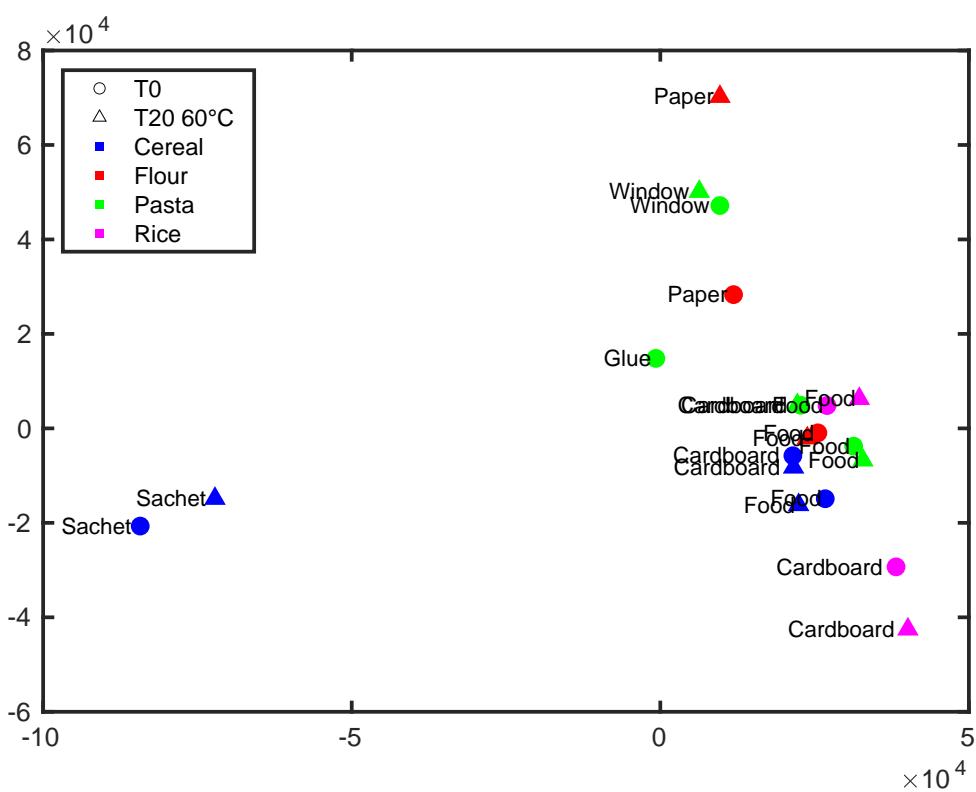


FIGURE 5.45 – Cartographie sur les dimensions 1×2 des emballages et aliments selon le critère de Levenshtein. Nous pouvons observer le déplacement des emballages avant et après vieillissement à 60°C pendant 20 jours.

Défi 2 : limites de détectabilité de NIAS fortuits

La méthode actuelle ne permet pas de fixer *a priori* un seuil à la détection de composés NIAS. La méthode est théoriquement invariante avec la concentration de la substance mais sa détectabilité notamment sur le critère dépend de l'échelle retenue, de la sensibilité et résolution du détecteur, de la complexité du mélange initial (en particulier de la proximité des pics et des forêts de pics dans les chromatogrammes).

Nous avons évoqué l'effet de loupe de la transformée en ondelettes (Figure 5.11) qui dépend de la largeur σ des pics en nombre de points et qui peut permettre en théorie de détecter des signaux d'amplitude isolés plus faible que les seuils considérés classiquement en chromatographie :

- limite de détection à un rapport signal sur bruit de 3
- limite de quantification à un rapport signal sur bruit de 10

Des détecteurs avec une fréquence d'acquisition et une résolution en masse plus élevées profiteraient *a priori* bien plus de cette méthodologie, car les signaux produits seraient mieux définis (plus larges en nombre de points) et les différents IC unitaires séparés (IC moins complexes).

Des acquisitions sur un petit nombre d'échantillons représentatifs de nos études ont été réalisées sur un appareil de GC-HRMS Orbitrap et ASAP-HRMS Orbitrap et seront traités dès que possible pour vérifier l'apport de la méthode aux techniques analytiques de pointe.

Défi 3 : combiner les approches développées avec les outils conventionnels de caractérisation des NIAS et des autres composés

Bien que les approches développées sous MATLAB puissent être portées sous de nombreux langages (R, Python, Rust) pour être intégrées dans d'autres projets open source, le travail dans l'espace transformé et le codage des signaux codages en lettres ne sont pas compatibles avec les outils conventionnels des fabricants de matériel et de logiciels de traitement. Le co-espace transformé présente de potentialités de déconvolution, filtrage et de comparaison qui justifient un retour à l'espace initial et la reconstruction de chromatogrammes avec un format de fichier standard netCDF ou mzXML. Il s'agit en particulier de pouvoir utiliser les bases de données de spectroscopie de masse pour identifier les composés soulignés par l'analyse.

Pour retourner à l'espace initial, il n'existe pas de transformée inverse pour l'ondelette de Ricker, car celle-ci n'est pas orthogonale. Nous avons alors proposé une méthode pseudo-inverse permettant de retrouver un signal proche du chromatogramme initial mais filtré des signaux issus de la colonne. Cette méthode se base sur la proposition de (Du et al., 2006) qui permet de déterminer la position et la largeur des pics dans un signal à l'aide de la transformée de Ricker appliquée à plusieurs échelles.

Nous avons discuté dans le manuscrit de l'existence d'une échelle optimale (3.4) pour la transformée propre à chaque pic pour laquelle l'amplification est maximale et permettrait la meilleure évaluation de l'aire. Nous avons ensuite construit un modèle reliant les aires de la transformée aux aires des gaussiennes initiales et illustré dans la Figure 5.46.

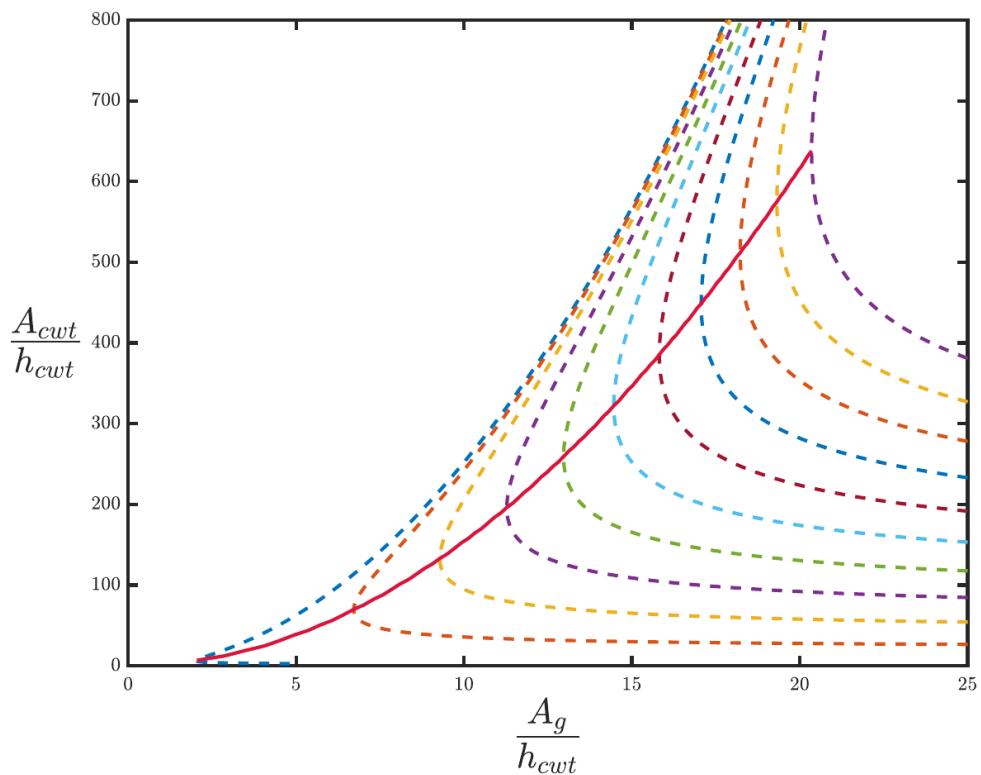


FIGURE 5.46 – Modèle de l'aire de la transformée à l'échelle optimale en fonction de l'aire de la gaussienne initiale. Les courbes en pointillés représentent l'évolution de l'aire de la transformée à différentes échelles pour une gaussienne de largeur fixe. La courbe continue en rouge constitue le modèle et relie les échelles optimales pour chaque largeur de gaussienne.

Il devient alors possible de reconstruire le signal chromatographique avec l'ensemble des signaux d'intérêt détectés par la méthode de [Du et al. \(2006\)](#) aux échelles optimales couplée au modèle d'aire que nous proposons en les assimilant à des gaussiennes parfaites. Cette transformée pseudo-inverse a cependant plusieurs limitations pouvant réduire sa fiabilité :

- les pics chromatographiques ont tendance à être asymétriques et non parfaitement gaussiens
- les faux positifs créent des pics qui n'existaient pas initialement, de même les faux négatifs suppriment des signaux d'intérêt
- la notion d'échelle optimale est difficile à définir lorsque plusieurs signaux d'intérêt se chevauchent ou fusionnent avec l'augmentation de l'échelle de la transformée.

Défi 4 : invariance entre appareils

Les alignements de chromatogrammes sont réalisés dans l'espace temporel. Ils devraient pouvoir être réalisés dans l'espace des indices Kovats. Parce que les indices de Kovats sont obtenus à partir des temps de rétention par une transformée monotone. Une partie du codage en segments monotones reste donc inchangé par passage en indices de Kovats. Les limites pour la comparaison d'échantillons mesurés sur des appareils et des protocoles distincts doivent être discutées.

Défi 5 : construction d'une base de données standardisées et échantillons de référence

L'espace chimique étudié est potentiellement immense et doit être développé via des points de référence connus qui posent les bases de la métrique du critère retenu. Cela inclut la définition et le positionnement de matériaux de référence et de profils de contamination. La possibilité de construire une base de données ouverte collaborative doit être discutée.

6 - Conclusion et perspectives

6.1 . Conclusions générales

6.1.1 . La reformulation du problème d'identification et de gestion des NIAS

Les travaux de thèse ont pour champ d'investigation la problématique de substances non intentionnellement ajoutées (NIAS) dans les matériaux au contact des aliments. Le concept de NIAS n'est en soi pas un terme scientifique, car il sous-tend une intention qui n'est pas démontrable par celui ou celle qui fait la mesure. Ce terme introduit et défendu par l'industrie dans une logique d'une décharge de responsabilité a remplacé progressivement dans la littérature les notions d'impuretés, de contaminations croisées, de composés néoformés ou de dégradation, qui qualifiaient le mécanisme physique ou chimique qui était à leur origine. Toutes ces notions ont toutefois un point commun, elles reposent sur une représentation parcimonieuse des contaminants : les substances, que l'on ne tracerait pas, seraient peu nombreuses et apparaîtraient de manière fortuite. Au démarrage de ces travaux, ce point de vue dominait. Les méthodes numériques de fouille des données de chimie analytique qui sont au cœur de ces travaux étaient envisagées comme une aide à l'identification de substances rares. L'identification même approximative de la substance couplée à une semi-quantification devait alimenter des outils du calcul de la migration dans les conditions d'utilisation envisagées (type d'aliment, temps et température). Il s'agissait ainsi de démontrer si l'exposition à cette substance serait acceptable en la comparant notamment aux seuils de préoccupation toxicologique défendus par l'EFSA. Les promulgations des directives (EU) 2019/904 et surtout la loi 2020-105 relative à la lutte contre le gaspillage et à l'économie circulaire ont profondément affecté notre compréhension de la thématique NIAS. Les matériaux de demain ne seraient pas établis sur la base de composés initiaux (matrices, résines, additifs) dont la pureté serait maîtrisée et où les NIAS ne seraient que des aléas maîtrisables par les bonnes pratiques de production. En prenant l'exemple des papiers et cartons recyclés, les NIAS sont ubiquitaires et signent la présence de matière recyclée, son origine, la nature et le nombre de recyclages. L'horizon chimique des substances considérées au sein du laboratoire national de référence des matériaux au contact des aliments était repoussé de manière très importante et ne se limitait plus aux 850 substances de la liste positive des matières plastiques, à la cinquantaine de substances de la fiche n°4 de la DGCCRF sur les matériaux au contact des aliments. L'appel à manifestation d'intérêt de l'Agence National de la Recherche du 19 septembre 2022 sur les nouvelles contraintes d'alimentarité des emballages et contenants alimentaires confirme cette analyse. Les NIAS ne doivent pas faire l'objet d'un traitement marginal, ils deviennent potentiellement les contaminants principaux dont on ne sait pas grand-chose aujourd'hui parce que non seulement notre compréhension n'est que partielle, les méthodes de caractérisation manquent et surtout parce que les nouvelles filières associées à l'économie circulaire ne sont pas encore en place. Pendant les deux dernières années, tout à chacun a pu observer le brunissement des bouteilles de boissons (eau, sodas, jus) du fait de la généralisation de l'utilisation du PET recyclé, plus dégradé et contenant des azurants issus des mélanges de la matière recyclée.

6.1.2 . L'introduction de notions thermodynamiques

Les constats précédents ont conduit à une reformulation des objectifs et des approches envisagés pour ces travaux. Les NIAS doivent être abordés de manière holistique comme un mélange et non comme des présences ponctuelles et indépendamment de la disponibilité des méthodes analytiques pour les détecter et les quantifier de manière individuelle. Les travaux ont visé ainsi à jeter les bases mathématiques de méthodes permettant de construire une empreinte d'un matériau inconnu et de réaliser des opérations mathématiques du type comparaison, soustraction ou filiation d'empreintes, corrélation d'empreintes avec des effets explicables de type formulation, procédé, origine, mode de collecte, de tri, de décontamination et de recyclage. Un fil d'Ariane a guidé les choix retenus : il y aurait une analogie entre l'information chimique et les mécanismes responsables des signatures observées (mélanges, transferts croisés, réactions chimiques) qui sont essentiellement sous contrôle entropique. L'application de concepts thermodynamiques à des mesures de chimie analytique permet de garantir une certaine exhaustivité des interprétations : aucune substance n'est ignorée parce qu'elles sont toutes comptabilisées, qu'elles soient séparables ou non, quantifiables ou non. Par ailleurs, il est apparu intéressant d'introduire un regard différent sur l'économie circulaire qui met essentiellement en avant des bénéfices liés au premier principe thermodynamique (réduction des coûts énergétiques et matière), mais ne discute pas des conséquences négatives liées à la dispersion des contaminants indépendamment des règles actuelles qui gèrent leur utilisation (liste positive ou négative, restrictions d'emploi, critères de pureté). Les concepts de type entropie offrent ainsi une opportunité de mesurer des dérives qui seraient liées à l'économie circulaire en fonction de la filière considérée. Les concepts ont été développés uniquement sur un spectromètre de masse de type simple quadrupôle couplé à la chromatographie en phase gazeuse avec un programme de température permettant de cibler simultanément les substances volatiles et semi-volatiles. La technique est relativement universelle et est partagée tant chez les fabricants, les utilisateurs (transformateurs, industriels de l'agroalimentaire), les recycleurs, les centres techniques et les autorités. Il serait ainsi théoriquement possible de faire fonctionner tous les acteurs en mode coopératif de manière à établir une représentation du flux de substances traversant la filière.

6.1.3 . Le codage du signal chromatographique et son interprétation

Le signal mesuré est au mieux une grandeur mathématique (vectorielle ou matricielle), mais ne traduit pas directement la complexité du mélange des substances dans un extrait du matériau obtenu par extraction Soxhlet ou par macération. Le signal du détecteur de masse comprend un bruit thermique issu des composés de dégradation et des contaminations de la colonne, et les composantes liées aux substances de l'extrait (solvants, étalons, IAS et NIAS). Le bruit thermique a été modélisé pour chaque ion (IC) comme un bruit de Poisson qui dépend du temps d'élution et du programme de température. Les signaux supérieurs à un critère de rejet statistique sont inclus dans l'empreinte. Afin de capturer également la signature de l'empreinte de substances présentes à faibles concentrations, un critère de rejet lâche basé sur la loi de Bienaymé-Tchebychev a été retenu. Même dans les zones très bruitées, il préserve jusqu'à 5% de l'intensité du signal. Le signal ainsi obtenu est codé après transformation avec une ondelette de Ricker à une échelle optimale qui se comporte comme un opérateur de dérivation à l'ordre 2 couplé à un filtre gaussien de même largeur que l'ondelette. Le signal est compressé en ne retenant que le signe de variation, la largeur et l'amplitude des segments monotones. Les ICs sont transformés en autant de séquences de 7 lettres (3 pour les segments croissants, 3 pour les décroissants, 1 pour les segments constants). Il est important de noter que bien que la transformée ne soit pas inversible, le signal initial peut être reconstruit en combinant les analyses en ondelettes aux échelles 22, 23 (échelle optimale moyenne), 24, 25. Un algorithme est proposé de manière à rendre compatibles les approches développées avec les outils chromatographiques (identification de substances avec les bases de données de spectre de masse) et de chimiométrie existants. La transformée en elle-même fournit de nombreux avantages. Elle supprime tous les signaux stationnaires issus de la colonne ou du détecteur (dérive temporelle, bruits stochastiques stationnaires) et amplifie tous les signaux dont la durée de vie coïncide avec la largeur de l'ondelette. Avec une échelle optimale de longueur 8, l'amplification est au plus d'un facteur 8 avec des gains compris entre 2 et 4 mesurés dans des conditions réelles. L'invariance en intensité et largeur est un point fort de la méthode qui permet de comparer des empreintes obtenues avec des variations sur les longueurs des colonnes et des températures.

6.1.4 . La comparaison de codes « empreintes » entre échantillons

La méthode du traitement de signal proposée permet sans calibration préalable de sortir les ions les plus significatifs pour un échantillon inconnu et sans référence connue. Bien que l'approche apporte une aide par rapport aux méthodes d'analyse des chromatogrammes utilisées par le LNE, la plus-value du travail réside dans la comparaison d'une empreinte complète (sans censure) d'un échantillon inconnu avec une référence acceptable. Cette possibilité remplit un besoin de l'industrie dans le cadre d'une économie linéaire et d'une économie circulaire : à partir de quand dévie un échantillon des standards de production ou de recyclage ? La réponse comprend deux niveaux. Quelle est l'amplitude de la différence ? Quelles parties du chromatogramme l'explique-t-elle ? Contrairement aux techniques classiques basées sur une quantification des signaux, la théorie de l'information ne requiert aucun préalable sur la linéarité de la réponse du détecteur et de la chaîne de mesure. Il n'y a toutefois pas de règle absolue pour mesurer la différence d'information contenue dans deux signaux. Nous avons construit deux critères suivant l'interprétation classique de l'information, soit sur la longueur d'édition, soit sur l'entropie en excès présente dans les deux signaux. La première interprétation mesure le nombre de changements qui doit être réalisé dans le signal pour transformer le code en un autre code. La seconde interprétation considère que chaque signal est en fait un échantillon d'une population de signaux et on cherche à estimer la quantité d'information qui n'est ni dans l'un ni dans l'autre. Aucune des deux méthodes n'établit de prévalence d'un signal sur l'autre et la distance obtenue est symétrique et obtenue par addition sur tous les IC (extensivité du critère). L'originalité de l'approche réside essentiellement dans un alignement des codes empreintes préalablement à leur comparaison. Cette approche permet d'établir une statistique robuste sur les parties communes à partir seulement de deux échantillons. La partie variable entre les deux empreintes peut contenir encore une entropie thermique résiduelle aléatoire. Les différences sont toutefois plus grandes quand les signaux sont issus d'une série de production ou de pratiques différentes. La comparaison de deux échantillons n'est pas limitée à des échantillons présentant une certaine homologie. On peut comparer, par exemple, un emballage et un aliment. S'ils n'ont pas été mis en contact, la partie commune doit être négligeable et les distances seront très grandes. Si l'emballage a été contact avec l'aliment et si des transferts ont eu lieu, les distances seront plus faibles et témoigneront d'un échange mutuel de substances. Le même procédé peut être utilisé pour détecter des contaminations croisées.

6.1.5 . Le concept d'espace chimique et de déplacement chimique dans cet espace

Les voisinages entre échantillons désignent des boules dans un espace métrique qui n'est pas connu a priori. Nous nous sommes posés des questions sur l'existence d'espaces chimiques qui permettraient de positionner des échantillons et d'interpréter leur éventuelle signification. Nous avons procédé avec prudence et choisi des explorations progressives soit par matrice (polyoléfine, polyester), soit par type d'économie (linéaire ou circulaire). L'existence d'un espace chimique a été testée d'abord sur un site industriel de production de polypropylène associé à quatre références, différentes formulations et conditions de mise en forme. L'itinéraire technologique comprend 4 étapes, les différences éventuelles sont essentiellement endogènes ou liées à la formulation. Parce que la préparation des extraits peut entraîner des variations faibles, mais détectables sur les profils d'extraction des substances, la boule minimale a été établie pour avoir la même taille que la plus grande variabilité entre échantillons. Ce qui est à l'extérieur de cette boule (espace nul) développe un espace chimique que le bruit ne peut pas expliquer. Avec une trentaine d'échantillons, un espace jusqu'à 7 dimensions a été reconstruit après un traitement en coordonnées principales. Ce qui nous a surpris, c'est que cet espace ne décrivait pas une variation infime des différences (ce à quoi on aurait pu s'attendre si les NIAS avait une origine fortuite et rare), mais 20 à 40% de la variance totale entre échantillons. Les effets de l'itinéraire technologique avec les effets de chaque étape qui apparaisse comme des déplacements chimiques. Le même effet entraîne le même vecteur de déplacement, quel que soit l'échantillon considéré. Sur les échantillons étudiés, les deux critères sont peu perturbés par la présence de colorants ou de charges. Les différences observées ne correspondent pas à la présence d'une ou deux substances, mais de plusieurs. Dans le détail, des variations (additions, disparitions) du profil oligomérique sont détectées. Il s'agit d'une chimie complexe qui peut être traduite sous la forme de déplacements chimiques. L'enregistrement d'empreintes à différents stades de la transformation permet de suivre une série industrielle et notamment le niveau de transformation de la matière du stade de granulé jusqu'au produit fini. L'analyse de la variance des effets imbriqués a permis d'identifier les dimensions de l'espace chimique qui expliquent au mieux la combinaison des effets. Les distances estimées dans l'espace développé par ces dimensions permettent de reconstruire la filiation complète des échantillons. Dans un espace défini par un petit nombre de dimensions (ex. 3), quatre distances avec quatre substances de référence choisies pour leur représentativité des transformations sont nécessaires pour positionner un échantillon inconnu d'une matière similaire à celle étudiée (ici un polypropylène isotactique).

6.1.6 . L'espace chimique dans le cadre de l'économie circulaire

L'étude contrôlée des effets des procédés et des formulations a montré que les dimensions n'ont pas une interprétation strictement causale. Aucune dimension ne décrit par exemple l'effet de la température seul, mais elles développent collectivement un espace dans lequel l'information s'accumule au fur et à mesure que le matériau évolue. En comparaison, le concept d'espace chimique dans le cadre d'une économie circulaire est beaucoup plus hypothétique, car il est associé à très grand nombre d'effets. La réductibilité de cet espace n'est pas testable directement, car il n'est pas possible d'agir sur la filière pour un ou plusieurs cycles de recyclage. Certaines étapes sont réalisables en laboratoire, mais il serait très difficile de reproduire les interactions entre flux de matériaux qui accompagnent leur collecte, lavage, recyclage et réutilisation. Parce que la disponibilité de la matière recyclée est régie par les réglementations en place et les règles du marché, un échantillonnage géographique a été réalisé à l'échelle de quatre continents (Europe, Amérique du Nord et du Sud, Asie, Australie) pour deux matrices polymères : le polypropylène (rPP) et le polyéthylène téréphthalate (rPET). Le rPP n'est disponible que pour le marché nord-américain et quatre groupes de rPP ont pu être collectés pour un total de six échantillons indépendants. Le rPET est disponible plus globalement et 25 gisements ont été identifiés et échantillonés pour un total de 136 échantillons. Avant ces travaux, il n'existe ni au LNE ni dans les autres laboratoires de référence en Europe de méthode pour comparer de manière systématiquement des gisements deux à deux. Malgré le petit nombre d'échantillons de rPP, l'espace développé est associé à des distances dix fois plus grandes que celles mesurées pour la mise en forme de PP vierges. Deux dimensions permettent de positionner de manière satisfaisante les échantillons en fonction de leur origine et mode de tri. Le concept de trajectoire et de directionnalité des déplacements chimiques en fonction de la sévérité des risques de contaminations croisées a été testé en reconstruisant la classification hiérarchique supposée des échantillons sur la base des seules mesures de distance. Sur les échantillons de rPET, une validation croisée complète a pu être réalisée en projetant de manière aléatoire un tiers des échantillons dans l'espace du tiers restant. L'opération répétée un millier de fois a permis d'estimer les incertitudes associées à la mesure de distances dans l'espace chimique du rPET. Quel que soit le critère retenu, les distances sont dix fois supérieures à celles observées pour le rPP. Le résultat était d'autant plus inattendu que les variations des signaux chromatographiques sont moins visibles dans le rPET : matrice vitreuse moins contaminée avant recyclage, pas de forêt de pics associée aux oligomères, plus faible formulation. Un classement hiérarchique des échantillons a été établi. Une interprétation simplifiée a été réalisée dans la logique de la réglementation européenne : la matière recyclée doit être aussi sûre que la matière vierge (règlement (EU) n° 2022/1616). La comparaison à la matière vierge permet d'établir trois groupes. Un groupe central de rPET très homogène, mais distant du PET vierge et deux groupes présentant une variabilité importante. Le premier groupe est intermédiaire entre le PET vierge et le rPET du groupe central et que l'industrie destine à l'eau embouteillée. À l'autre extrémité, on trouve un groupe de rPET très hétérogène, distinct des autres PET et qui rassemble les rPET non décontaminés et/ou issus de processus non sélectifs. Les deux critères permettent d'identifier ces rPET extrêmes qui ne sont pas acceptables sur le marché européen. De manière pratique, le critère entropique tend à donner en aveugle de meilleures performances pour la discrimination et la hiérarchisation des échantillons. De manière surprenante, les rPET recyclés chimiquement sont distincts des PET vierges et présentent plus de proximité avec le groupe extrême incluant les gisements non conformes pour le contact alimentaire.

6.2 . Perspectives

6.2.1 . Le nombre de l'économie circulaire : \daleth

Dix à vingt dimensions décrivent l'économie circulaire du rPET. La traçabilité insuffisante des échantillons ne nous a pas permis de comprendre pourquoi autant de déplacements chimiques indépendants sont nécessaires à leur description, mais nous proposons une interprétation physique de l'économie circulaire de cet espace. La matière recyclée traverse au cours de sa vie cet espace le long d'une trajectoire qui est déterminée par les substances gagnées ou perdues et que nous mesurons comme le flux cumulé échangé le long de cette trajectoire. La distance totale est associée à un nombre sans dimension \daleth qui caractériserait le bilan net des substances échangées. Les directions explorées dans l'espace de l'économie circulaire seraient la seule conséquence du profil de contamination. En considérant que vingt à trente lettres codes sont nécessaires pour décrire une substance dans nos approches, l'espace de contamination du rPET est caractérisé par des milliers des substances bien au-delà des capacités disponibles pour leur identification. Nous pensons que comme l'entropie est croissante pour un système fermé, la variation du nombre \daleth tendrait à être monotone avec le nombre de recyclages et il pourrait être compris comme une longueur curviligne le long du chemin. La distance euclidienne mesurerait seulement une partie de cet effet. La distance curviligne mesurerait l'âge réel moyen d'un gisement et le nombre de recyclages associés. La décontamination permettrait de courber la trajectoire pour la rapprocher d'un profil moyen caractérisé par les molécules les moins volatiles et qui ne sont pas éliminées lors du recyclage. Les propriétés cycliques de \daleth pourraient être étudiées en comparant l'évolution des empreintes chimiques à différents stades de l'étape de recyclage : lavage, décontamination, granulation... Les étapes de décontamination n'éliminent pas les substances, elles en réduisent les concentrations. L'hypothèse d'une réduction de distances à la matière vierge lors recyclage du fait des limites de détection de la méthode a été testée par ajout dosé de cinq substances en combinaison (résultats non montrés). Il a été ainsi montré que les échantillons qui contenaient tous les mêmes substances étaient vus différemment quand une ou plusieurs espèces devenaient non détectables. D'autres effets comme la coélution des substances affectent la mesure des distances.

6.2.2 . L'hypothèse de détectabilité des transferts via l'échange des informations

La méthode développée a été testée sur des aliments secs et les différentes parties de l'emballage en contact dans le cadre du projet ANR FoodSafeBioPack (référence ANR-20-CE21-0009, 2021-2024). La méthode des distances proposée montre que les transferts croisés entre un aliment et son emballage s'accompagnent de déplacements chimiques opposés vers un point hypothétique qui correspondrait à leur mélange. Il serait ainsi possible de reconstruire la causalité des contaminations et d'écartier les parties de l'emballage qui ne seraient pas incriminées. Des comparaisons similaires ont été réalisées pour les plastiques agricoles (résultats non montrés). Des mélanges numériques des chromatogrammes (résultats non montrés) ont confirmé la convergence, mais contrairement à l'intuition les trajectoires ne sont pas linéaires, mais courbes. La non-linéarité des lois de mélange peut être vue comme une déviation de l'extensivité des critères proposée ou comme la conséquence de l'effet de censure appliquée à l'information par les filtres numériques.

6.2.3 . Implémentation et utilisation de la méthode dans les plans de surveillance de la matière recyclée

La méthode complète est disponible sous la forme d'un ensemble de fonctions développées en langage Matlab et déposées sur le site GitHub. La documentation est cours d'écriture. La méthode est en cours de consolidation pour une utilisation en production au LNE. La méthode ne requiert pas d'actions particulières de la part de l'opérateur autre que l'exportation d'un fichier XML-MZ. La transformée inverse est disponible sous la forme d'un fichier XML-MZ qui peut ne contenir que les différences entre deux ou plusieurs échantillons. Les composés peuvent être identifiés avec les bases de données de spectre de masse, mais deux difficultés compliquent cette tâche : leur nombre et la présence d'IC peu ou pas résolus. La comparaison des identifications avec des techniques de spectroscopie de masse à haute résolution (ASAP MS, GC-MS Orbitrap) a confirmé la complexité des extraits en nombre de substances sur quelques échantillons (résultats non montrés). Cet effort doit être encore poursuivi. La construction d'un espace chimique dans lequel les déplacements chimiques pourraient être caractérisés dans le cadre d'une économie circulaire est dépendante de la disponibilité d'échantillons représentatifs qui définissent des distances et des orientations de référence. Il est urgent de construire une échantillothèque représentative à l'échelle du marché international (des milliers ou dizaines de milliers). Les échantillons et leurs extraits devront être stockés (à froid) pour plusieurs années afin d'alimenter les développements futurs.

6.2.4 . Le partage de l'information chimique et la coopération des acteurs

Les codes des empreintes peuvent être aisément stockés sur un serveur avec une référence anonyme et permettre ainsi leurs comparaisons avec des futurs échantillons. Le format de fichier peut aisément être standardisé puisqu'il s'agit essentiellement d'une séquence ASCII. L'invariance avec le type de colonne, de détecteur et de programme de température n'est pas assurée et une comparaison interlaboratoire est requise pour établir un mode opératoire minimal qui garantit la robustesse du calcul des distances. Sur cette base, on pourrait imaginer que l'information de type empreinte chimique (ou ses coordonnées dans l'espace chimique) accompagne le produit tout au cours de son cycle de vie et oriente les valorisations futures. La technologie blockchain permettrait d'agréger l'information de manière sûre. À très court terme, la méthode pourrait être déployée pour prévenir sur les sites de recyclage le risque de upcycling de la matière collectée qui conduirait à mettre sur le marché des gisements impropre au contact alimentaire. La même technique de triage pourrait être appliquée à toute matière (plastique, cellulosique) pour décider si elle doit être utilisée derrière une barrière fonctionnelle ou non.

Bibliographie

- ADEME. Cartographie des couples produit / type et matériaux d'emballages : Cartographie et état des lieux. 2021. URL <https://librairie.ademe.fr/dechets-economie-circulaire/5272-cartographie-des-couples-produit-type-et-materiaux-d-emballages.html>. (date accessed 17/12/2022).
- ADEME. Chiffres clés. 2022. URL <https://expertises.ademe.fr/economie-circulaire/dechets/chiffres-cles-observation/chiffres-cles>. (date accessed 15/12/2022).
- D. An, K. A. Meissner, P. Nurowski, and R. Penrose. Apparent evidence for Hawking points in the CMB Sky . *Monthly Notices of the Royal Astronomical Society*, 495(3) :3403–3408, 05 2020. ISSN 0035-8711. doi : 10.1093/mnras/staa1343. URL <https://doi.org/10.1093/mnras/staa1343>.
- ANR. Emballages et contenants alimentaires et politiques de durabilité : nouvelles contraintes d'alimentarité et impacts sur la nutrition et la santé – appel à manifestation d'intérêt, 2022. URL <https://anr.fr/fr/france-2030/france2030/call/emballages-et-contenants-alimentaires-et-politiques-de-durabilite-nouvelles-contraintes-dalimenta/>.
- V. P. Balema, I. Z. Hlova, S. L. Carnahan, M. Seyedi, O. Dolotko, A. J. Rossini, and I. Luzinov. Depolymerization of polystyrene under ambient conditions. *New Journal of Chemistry*, 45(6) :2935–2938, 2021. ISSN 1144-0546. doi : 10.1039/DONJ05984F. URL <http://dx.doi.org/10.1039/DONJ05984F>.
- A. Baner, R. Franz, and O. Piringer. Alternative methods for the determination and evaluation of migration potential from polymeric food contact materials. *Deutsche Lebensmittel-Rundschau*, 90(6) :181–185, 1994. URL https://serials.unibo.it/cgi-ser/start/en/spogli/df-s.tcl?prog_art=3884745&language=ENGLISH&view=articoli.
- A. Baner, J. Brandsch, R. Franz, and O. Piringer. The application of a predictive migration model for evaluating the compliance of plastic materials with european food regulations. *Food Additives and Contaminants*, 13(5) :587–601, 1996. ISSN 0265-203X. doi : 10.1080/02652039609374443. URL <https://doi.org/10.1080/02652039609374443>.
- E. Barthelemy, D. Spyropoulos, M. R. Milana, K. Pfaff, N. Gontard, E. Lampi, and L. Castle. Safety evaluation of mechanical recycling processes used to produce polyethylene terephthalate (pet) intended for food contact applications. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*, 31(3) :490–7, 2014. ISSN 1944-0057 (Electronic) 1944-0057 (Linking). doi : 10.1080/19440049.2013.871755. URL <https://www.ncbi.nlm.nih.gov/pubmed/24341373>.
- A. Bayat. Bioinformatics. *BMJ*, 324(7344) :1018–1022, 2002. ISSN 0959-8138. doi : 10.1136/bmj.324.7344.1018. URL <https://www.bmjjournals.org/content/324/7344/1018>.
- M. Biedermann, J.-E. Ingenhoff, M. Barbanera, D. Garbini, and K. Grob. Migration of mineral oil into noodles from recycled fibres in the paperboard box and the corrugated board transport box as well as from printing inks : A case study. *Packaging Technology and Science*, 24(5) :281–290, 2011. ISSN 0894-3214. doi : 10.1002/pts.937. URL <https://doi.org/10.1002/pts.937>.

- M. Biedermann, J.-E. Ingenhoff, G. Dima, M. Zurfluh, S. Biedermann-Brem, L. Richter, T. Simat, A. Harling, and K. Grob. Migration of mineral oil from printed paperboard into dry foods : survey of the german market. part ii : advancement of migration during storage. *European Food Research and Technology*, 236(3) :459–472, 2013. ISSN 1438-2377. doi : 10.1007/s00217-012-1909-2. URL <https://doi.org/10.1007/s00217-012-1909-2>. Times Cited : 5.
- Y. Bruinen de Bruin, A. Franco, A. Ahrens, A. Morris, H. Verhagen, S. Kephalopoulos, V. Dulio, J. Slobodnik, D. T. H. M. Sijm, T. Vermeire, T. Ito, K. Takaki, J. De Mello, J. Bessems, M. Zare Jeddi, C. Tanarro Gozalo, K. Pollard, J. McCourt, and P. Fantke. Enhancing the use of exposure science across eu chemical policies as part of the european exposure science strategy 2020–2030. *Journal of Exposure Science and Environmental Epidemiology*, 32(4) :513–525, 2022. ISSN 1559-064X. doi : 10.1038/s41370-021-00388-4. URL <https://doi.org/10.1038/s41370-021-00388-4>.
- A. Cifuentes. Food analysis and foodomics. *Journal of Chromatography A*, 1216(43) :7109, 2009. ISSN 0021-9673. doi : <https://doi.org/10.1016/j.chroma.2009.09.018>. URL <https://www.sciencedirect.com/science/article/pii/S0021967309013545>.
- A. Datta, B. Nicolaï, O. Vitrac, P. Verboven, F. Erdogan, F. Marra, F. Sarghini, and C. Koh. Computer-aided food engineering. *Nature Food*, 3(11) :894–904, 2022. ISSN 2662-1355. doi : 10.1038/s43016-022-00617-5. URL <https://doi.org/10.1038/s43016-022-00617-5>.
- DGCCRF. Fiche mcda n°4 (v02 – 01/01/2019) aptitude au contact alimentaire des matériaux organiques à base de fibres végétales destinés à entrer en contact avec des denrées alimentaires, 2019. URL <https://www.economie.gouv.fr/dgccrf/materiaux-organiques-a-base-fibres-vegetales>.
- Direction Générale de la Santé. Publication du bilan à mi-parcours de la deuxième stratégie nationale sur les perturbateurs endocriniens. 2022. URL <https://solidarites-sante.gouv.fr/sante-et-environnement/risques-microbiologiques-physiques-et-chimiques/article/perturbateurs-endocriniens>.
- P. Du, W. A. Kibbe, and S. M. Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22 (17) :2059–2065, jul 2006. doi : 10.1093/bioinformatics/btl355. URL <https://doi.org/10.1093/bioinformatics/btl355>.
- EC. Commission regulation (eu) no 10/2011 of 14 january 2011 on plastic materials and articles intended to come into contact with food text with eea relevance. *Official Journal of the European Union*, L12 :1–89, 2011. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011R0010>.
- EC. Webinar on the evaluation and revision of the eu rules on food contact materials, 2021a. URL https://ec.europa.eu/food/system/files/2021-01/cs_fcm_eval-workshop_20210120_pres.pdf.
- EC. Proposal for a regulation of the european parliament and of the council on shipments of waste and amending regulations (eu) no 1257/2013 and (eu) no 2020/1056. 2021b. URL <https://eur-lex.europa.eu>.

EC. Call for evidence : Environmental impact of waste management – revision of EU waste framework, 2022. URL https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13225-Environmental-impact-of-waste-management-revision-of-EU-waste-framework_en.

EFSA. Scientific opinion on mineral oil hydrocarbons in food. *EFSA Journal*, 10(6) :2704, 2012. doi : 10.2903/j.efsa.2012.2704. URL <https://doi.org/10.2903/j.efsa.2012.2704>.

ESA. Outgassing database, 2018. URL http://esmat.esa.int/services/outgassing_data/outgassing_data.html.

X. Fang and O. Vitrac. Predicting diffusion coefficients of chemicals in and through packaging materials. *Critical Reviews in Food Science and Nutrition*, 57(2) :275–312, 2017. doi : 10.1080/10408398.2013.849654. URL <https://doi.org/10.1080/10408398.2013.849654>.

P. Fantke, C. Cinquemani, P. Yaseneva, J. De Mello, H. Schwabe, B. Ebeling, and A. A. Lapkin. Transition to sustainable chemistry through digitalization. *Chem*, 7(11) :2866–2882, 2021. ISSN 2451-9294. doi : 10.1016/j.chempr.2021.09.012. URL <https://doi.org/10.1016/j.chempr.2021.09.012>.

FAO. *Assuring Food Safety and Quality. Guidelines for Strengthening National Food Control Systems*. FAO Food and Nutrition Paper. FAO, 2003. URL <https://www.fao.org/sustainable-food-value-chains/library/details/en/c/265857/>.

T. Foltýnek, N. Meuschke, and B. Gipp. Academic plagiarism detection : A systematic literature review. *ACM Comput. Surv.*, 52(6), oct 2019. ISSN 0360-0300. doi : 10.1145/3345317. URL <https://doi.org/10.1145/3345317>.

Food Packaging Forum. Dossier – mineral oil hydrocarbons. Report, Food Packaging Forum, 2017. URL https://www.foodpackagingforum.org/fpf-2016/wp-content/uploads/2017/06/FPF_Dossier12_mineral-oil-hydrocarbons.pdf.

FoodWatch. Mineral oil in food – results of the foodwatch test. Report, FoodWatch, 2015. URL https://www.foodwatch.org/fileadmin/foodwatch_international/reports/2015-Mineral_oils_in_food.pdf.

FoodWatch. Contamination des aliments par les huiles minérales : résultat des tests foodwatch. Report, FoodWatch, 2021. URL https://www.foodwatch.org/fileadmin/-FR/Campagnes/Alimentation_et_sante/Huiles_minerales/RAPPORT_2021_foodwatch_huiles_minerales.pdf.

R. Franz and F. Welle. Contamination levels in recollected pet bottles from non-food applications and their impact on the safety of recycled pet for food contact. *Molecules*, 25 (21), 2020. ISSN 1420-3049. doi : 10.3390/molecules25214998. URL <https://www.mdpi.com/1420-3049/25/21/4998>.

T. P. Gandek, T. A. Hatton, and R. C. Reid. Batch extraction with reaction : phenolic antioxidant migration from polyolefins to water. 1. theory. *Industrial and Engineering Chemistry Research*, 28(7) :1030–1036, 1989a. ISSN 0888-5885. doi : 10.1021/ie00091a023. URL <https://doi.org/10.1021/ie00091a023>.

- T. P. Gandek, T. A. Hatton, and R. C. Reid. Batch extraction with reaction : phenolic antioxidant migration from polyolefins to water. 2. experimental results and discussion. *Industrial and Engineering Chemistry Research*, 28(7) :1036–1045, 1989b. ISSN 0888-5885. doi : 10.1021/ie00091a024. URL <https://doi.org/10.1021/ie00091a024>.
- G. Gillet, O. Vitrac, D. Tissier, P. Saillard, and S. Desobry. Development of decision tools to assess migration from plastic materials in contact with food. *Food Additives and Contaminants Part a-Chemistry Analysis Control Exposure and Risk Assessment*, 26(12) : 1556–1573, 2009. ISSN 0265-203X. doi : 10.1080/19440040903271355. URL <https://doi.org/10.1080/19440040903271355>.
- G. Gillet, O. Vitrac, and S. Desobry. A fast method to assess the composition of a polyolefin : An application to compliance testing of food contact materials. *Journal of Applied Polymer Science*, 119(3) :1492–1515, 2011. doi : <https://doi.org/10.1002/app.32950>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/app.32950>.
- M. Hilton, C. Sherrington, A. Massie, J. Papineschi, P. Jones, and D. Hogg. *Study to support preparation of the Commission’s guidance for extended producer responsibility scheme : final report*. European, Commission Directorate-General for Environment. Publications Office, 2020. doi : doi/10.2779/55232.
- K. ; Hinrichs and O. Piringer. Evaluation of migration models to be used under directive 90/128/eec ; final repot contract smt4-ct9867513. eur 20604 en. *European Commission, Directorate General for Research*, 2002.
- E. Hoekstra, R. Brandsch, C. Dequatre, P. Mercea, M. R. Milana, A. Stoermer, X. Trier, O. Vitrac, A. Schaefer, and C. Simoneau. Practical guidelines on the application of migration modelling for the estimation of specific migration. Report, European Commission, Joint Research Centre, Institute for Health and Consumers Protection, 2015. URL <https://ec.europa.eu/jrc/en/publication/practical-guidelines-application-migration-modelling-estimation-specific-migration>.
- A. Hyvärinen. Independent component analysis : recent advances. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 371(1984) : 20110534, 2013. doi : 10.1098/rsta.2011.0534. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2011.0534>.
- IARC. Styrene, styrene-7,8-oxide, and quinoline. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, 121, 2019.
- R. J. Jaeger and R. J. Rubin. Plasticizers from plastic devices extraction, metabolism, and accumulation by biological systems. *Science*, 170(3956) :460–2, 1970. ISSN 0036-8075 (Print) 0036-8075. doi : 10.1126/science.170.3956.460. URL <https://doi.org/10.1126/science.170.3956.460>.
- N. Jawerth. Comment la rt-pcr en temps réel permet-elle dedétecer le virus de la covid-19 ? *Bulletin de l’AIEA*, 61-2 :8–11, Juin 2020. URL <https://www.iaea.org/fr/bulletin/comment-la-rt-pcr-en-temps-reel-permet-elle-de-detecter-le-virus-de-la-covid-19>.

- E. Joltreau. Five years after china's plastic import ban. have europeans taken responsibility ? *IFRI*, 2022. URL https://www.ifri.org/sites/default/files/atoms/files/joltreau_plastic_waste_after_chinas_ban_2022.pdf.
- D. L. Jow and D. Scott. Re-evaluating evidence for hawking points in the cmb. *Journal of Cosmology and Astroparticle Physics*, 2020(03) :021, mar 2020. doi : 10.1088/1475-7516/2020/03/021. URL <https://dx.doi.org/10.1088/1475-7516/2020/03/021>.
- E. Karamfilova and M. Sacher. Food contact materials regulation (EC) 1935/2004 : European implementation assessment study, 2016. URL [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/581411/EPRS_STU\(2016\)581411_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/581411/EPRS_STU(2016)581411_EN.pdf).
- A. Kassouf, J. Maalouly, H. Chebib, D. N. Rutledge, and V. Ducruet. Chemometric tools to highlight non-intentionally added substances (NIAS) in polyethylene terephthalate (PET). *Talanta*, 115 :928–937, oct 2013. doi : 10.1016/j.talanta.2013.06.029. URL <https://doi.org/10.1016/j.talanta.2013.06.029>.
- J. Kermorvant, P. M. Nguyen, C. Lyathaud, S. Domenek, and O. Vitrac. Evaluation et gestion des flux de matières recyclées pour le contact alimentaire. In *Vieillissement des Polymères (VIP)*, Paris, France, June 2022a.
- J. Kermorvant, P. M. Nguyen, C. Lyathaud, S. Domenek, and O. Vitrac. Coding and tracking contaminants of food contact materials by wavelet analysis. In *FoodSim*, Ghent, Belgium, April 2022b.
- J. Kermorvant, P.-M. Nguyen, C. Lyathaud, S. Domenek, and O. Vitrac. Coding and tracking contaminants of food contact materials by wavelet analysis. *Food Research International*, 2022c.
- J. Kermorvant, P.-M. Nguyen, and O. Vitrac. L'économie circulaire et la sécurité sanitaire des emballages alimentaires : pourquoi elles s'opposent, pourquoi elles doivent aller de concert. *Cahiers de nutrition et de diététique*, 2023.
- V. I. Levenstein. Binary codes capable of correction deletions, insertions and reversals. *Soviet Physics Doklady*, 163(4) :845–848, 1965.
- W. Limm and H.C. Hollifield. Modelling of additive diffusion in polyolefins. *Food Additives and Contaminants*, 13(8) :949–967, 1996. ISSN 0265-203X. doi : 10.1080/02652039609374482. URL <https://doi.org/10.1080/02652039609374482>.
- R. Lorenzini, K. Fiselier, M. Biedermann, M. Barbanera, I. Braschi, and K. Grob. Saturated and aromatic mineral oil hydrocarbons from paperboard food packaging : estimation of long-term migration from contents in the paperboard and data on boxes from the market. *Food Additives and Contaminants : Part A*, 27(12) :1765–1774, 2010. ISSN 1944-0049. doi : 10.1080/19440049.2010.517568. URL <https://doi.org/10.1080/19440049.2010.517568>.
- R. Lorenzini, M. Biedermann, K. Grob, D. Garbini, M. Barbanera, and I. Braschi. Migration kinetics of mineral oil hydrocarbons from recycled paperboard to dry food : monitoring of two real cases. *Food Additives and Contaminants : Part A*, 30(4) :760–770, 2013. ISSN 1944-0049. doi : 10.1080/19440049.2013.766765. URL <https://doi.org/10.1080/19440049.2013.766765>.

Ministère de la Transition Ecologique. Quel potentiel 3r (réduction, réemploi, recyclage) d'ici 2025 ? pour les emballages en plastique, 2020. URL
https://www.ecologie.gouv.fr/sites/default/files/Rapport_Quel_Potentiel_3R.pdf.

Ministère de la Transition Ecologique. Stratégie nationale « 3r » pour les emballages en plastique à usage unique - annexe 6 : fiches sectorielles, 2022. URL
https://www.consultations-publiques.developpement-durable.gouv.fr/IMG/pdf/annexe_6_de_la_strategie_-_fiches_sectorielles.pdf.

J. Muncke. Exposure to endocrine disrupting compounds via the food chain : Is packaging a relevant source ? *Sci Total Environ*, 407(16) :4549–59, 2009. ISSN 0048-9697. doi : 10.1016/j.scitotenv.2009.05.006. 1879-1026 Muncke, Jane Journal Article Review Netherlands Sci Total Environ. 2009 Aug 1;407(16) :4549-59. doi : 10.1016/j.scitotenv.2009.05.006. Epub 2009 May 30.

P. E. S. Munekata, M. Pateiro, M. López-Pedrouso, M. Gagaoua, and J. M. Lorenzo. Foodomics in meat quality. *Current Opinion in Food Science*, 38 :79–85, 2021. ISSN 2214-7993. doi : <https://doi.org/10.1016/j.cofs.2020.10.003>. URL
<https://www.sciencedirect.com/science/article/pii/S2214799320300813>.

NASA. Outgassing data for selecting spacecraft materials, 2018. URL
<https://outgassing.nasa.gov/>.

National standard of the people's republic of China. Gb 9685-2016 : National food safety standard - standard for uses of additives in food contact materials and their products. 2016. URL <https://www.chinesestandard.net/PDF.aspx/GB9685-2016>.

S. Navlakha and Z. Bar-Joseph. Algorithms in nature : the convergence of systems biology and computational thinking. *Molecular Systems Biology*, 7(1) :546, 2011. doi : <https://doi.org/10.1038/msb.2011.78>. URL
<https://www.embopress.org/doi/abs/10.1038/msb.2011.78>.

P.-M. Nguyen, A. Goujon, P. Sauvegrain, and O. Vitrac. A computer-aided methodology to design safe food packaging and related systems. *AICHE Journal*, 59(4) :1183–1212, 2013. ISSN 1547-5905. doi : 10.1002/aic.14056. URL <http://dx.doi.org/10.1002/aic.14056>.

P.-M. Nguyen, C. Lyathaud, and O. Vitrac. A two-scale pursuit method for the tailored identification and quantification of unknown polymer additives and contaminants by proton nmr. *Industrial & Engineering Chemistry Research*, 54(10) :2667–2681, mar 2015. doi : 10.1021/ie503592z. URL <https://doi.org/10.1021/ie503592z>.

P.-M. Nguyen, J. M. Julien, C. Breysse, C. Lyathaud, J. Thébault, and O. Vitrac. Project safefoodpack design : case study “indirect migration from paper and boards”. *Food Additives and Contaminants : Part A*, 34(10) :1703–1720, 2017. doi : 10.1080/19440049.2017.1315777. URL
<http://dx.doi.org/10.1080/19440049.2017.1315777>.

P.-M. Nguyen, S. Dorey, and O. Vitrac. The ubiquitous issue of cross-mass transfer : Applications to single-use systems. *Molecules*, 24(19) :3467 ; <https://doi.org/10.3390/molecules24193467>, 2019. doi : 10.3390/molecules24193467. URL
<https://doi.org/10.3390/molecules24193467>.

P.-M. Nguyen, J. Kermorvant, C. Lyathaud, S. Domenek, and O. Vitrac. Nouvelles stratégies pour l'évaluation et la gestion des matériaux recyclés pour le contact alimentaire. In *SFGP*, Toulouse, France, November 2022.

NRC. *Risk assessment in the federal government : Managing the process*. National Academy Press, Washington, DC, 1983. doi : 10.17226/366. URL <http://dx.doi.org/10.17226/366>.

OECD. Environment ministers' commitments on plastics : National-level visions, actions and plans announced at the 2022 oecd council at ministerial level (mcm). 2022. URL <https://www.oecd.org/environment/ministerial/outcomes/Environment-Ministers-commitments-on-plastics.pdf>.

I. Ortea. Foodomics in health : Advanced techniques for studying the bioactive role of foods. *TrAC Trends in Analytical Chemistry*, 150 :116589, 2022. ISSN 0165-9936. doi : <https://doi.org/10.1016/j.trac.2022.116589>. URL <https://www.sciencedirect.com/science/article/pii/S0165993622000723>.

H.-J. Paek and T. Hove. Risk perceptions and risk characteristics, 2017-03-29 2017. URL <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-283>.

O. Penrose. *Foundations of Statistical Mechanics : A Deductive Treatment*. Pergamon : Oxford, UK, 1970.

République Française. Décret n° 2021-517 du 29 avril 2021 relatif aux objectifs de réduction, de réutilisation et de réemploi, et de recyclage des emballages en plastique à usage unique pour la période 2021-2025. *JORF*, 2021. URL <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043458675#:~:text=Un%20objectif%20de%20tendre%20vers,est%20fix%C3%A9%20%C3%A0%201%C3%A9ch%C3%A9ance>.

I. Severin, P.-M. Nguyen, A. Platel, C. Lyathaud, J. M. Julien, J. Kermorvant, H. Moche, O. Vitrac, S. Domenek, and M.-C Chagnon. Packsafe : Integrated chemical safety assessment of food contact articles (anr-21-ce21-0004 project : 2022-2025). In *7th International Packaging Materials Symposium (ILSI)*, Digital event, May 2022.

C. E. Shannon. Prediction and entropy of printed english. *The Bell System Technical Journal*, 30(1) :50–64, 1951. doi : 10.1002/j.1538-7305.1951.tb01366.x. URL <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>.

V. Silano, J. M. Barat Baviera, C. Bolognesi, A. Chesson, P. S. Cocconcelli, R. Crebelli, D. M. Gott, K. Grob, C. Lambré, E. Lampi, M. Mengelers, A. Mortensen, I.-L. Steffensen, C. Tlustos, H. Van Loveren, L. Vernis, H. Zorn, L. Castle, E. Di Consiglio, R. Franz, N. Hellwig, M. R. Milana, K. Pfaff, M. Carfi, E. Van Haver, and G. Rivière. Assessment of the impact of the iarc monograph vol. 121 on the safety of the substance styrene (fcm no 193) for its use in plastic food contact materials. *EFSA Journal*, 18(10) :e06247, 2020. ISSN 1831-4732. doi : <https://doi.org/10.2903/j.efsa.2020.6247>. URL <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6247>. (CEP).

A. Sturaro, G. Parvoli, R. Rella, S. Bardati, and L. Doretti. Food contamination by diisopropylnaphthalenes from cardboard packages. *International Journal of Food Science and Technology*, 29(5) :593–603, 1994. ISSN 0950-5423. doi :

- <https://doi.org/10.1111/j.1365-2621.1994.tb02102.x>. URL
<https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.1994.tb02102.x>.
- C. Tsallis. Beyond boltzmann–gibbs–shannon in physics and elsewhere. *Entropy*, 21(7), 2019. ISSN 1099-4300. doi : 10.3390/e21070696. URL
<https://www.mdpi.com/1099-4300/21/7/696>.
- O. Vitrac and M. Hayert. Risk assessment of migration from packaging materials into foodstuffs. *AIChe Journal*, 51(4) :1080–1095, 2005. ISSN 0001-1541. doi : 10.1002/aic.10462. URL <https://doi.org/10.1002/aic.10462>.
- O. Vitrac and M. Hayert. *Design of safe packaging materials under uncertainty*, book section 215-292. Nova Science Publishers, New York, 2007. ISBN 9781600214868. URL
<https://books.google.fr/books?id=iSbICeDnyDUC>.
- O. Vitrac and J.-C. Leblanc. Consumer exposure to substances in plastic packaging. i. assessment of the contribution of styrene from yogurt pots. *Food Additives and Contaminants*, 24(2), 2007. ISSN 0265-203X(print)|1464-5122(electronic). doi : 10.1080/02652030600888618. URL <https://doi.org/10.1080/02652030600888618>.
- O. Vitrac, B. Challe, J.-C. Leblanc, and A. Feigenbaum. Contamination of packaged food by substances migrating from a direct-contact plastic layer : Assessment using a generic quantitative household scale methodology. *Food Additives and Contaminants*, 24(1) :75–94, 2007. ISSN 0265-203X. doi : 10.1080/02652030600888550. URL
<http://dx.doi.org/10.1080/02652030600888550>.
- O. Vitrac, J. Kermorvant, H. Ayoub, N. Daoud, and P.-M. Nguyen. Le recyclage et le réemploi des emballages peuvent-ils entraîner un risque systémique de contamination des aliments ? *Industries Alimentaires et Agricoles*, Emballage Alimentaire :18–22, 2021.
- O. Vitrac, J. Kermorvant, H. Ayoub, and P. M. Nguyen. New strategies to evaluate and manage recycled materials for food contact : a european point of view. In *ACS Fall (Division of Agricultural and Food Chemistry)*, Chicago, USA, August 2022a.
- O. Vitrac, J. Kermorvant, and P. M. Nguyen. Computer-aided design of safe food packaging with maximized recycled content. In *7th International Packaging Materials Symposium (ILSI)*, Digital event, May 2022b.
- WHO. Exposome-explorer : First database dedicated to biomarkers of exposure to environmental risk factors for diseases., 2019. URL <http://exposome-explorer.iarc.fr>.
- X. Zeng and J. Li. Emerging anthropogenic circularity science : Principles, practices, and challenges. *iScience*, 24(3) :102237, 2021. ISSN 2589-0042. doi :
<https://doi.org/10.1016/j.isci.2021.102237>. URL
<https://www.sciencedirect.com/science/article/pii/S2589004221002054>.
- Y. Zhu, B. Guillemat, and O. Vitrac. Rational design of packaging : Toward safer and ecodesigned food packaging systems. *Frontiers in Chemistry*, 7(349), 2019a. ISSN 2296-2646. doi : 10.3389/fchem.2019.00349. URL
<https://www.frontiersin.org/article/10.3389/fchem.2019.00349>.

Y. Zhu, P.-M. Nguyen, and O. Vitrac. *Risk Assessment of Migration From Packaging Materials Into Food*, page 64. Elsevier, Amsterdam, 2019b. ISBN 978-0-08-100596-5. doi : 10.1016/B978-0-08-100596-5.22501-8. URL
<https://doi.org/10.1016/B978-0-08-100596-5.22501-8>.

Y. Zhu, F. Welle, and O. Vitrac. A blob model to parameterize polymer hole free volumes and solute diffusion. *Soft Matter*, 15 :8912–8932, 2019c. doi : 10.1039/C9SM01556F. URL
<http://dx.doi.org/10.1039/C9SM01556F>.