



HAL
open science

Identification des molécules responsables de l'odeur d'un mélange gazeux – Vers une automatisation de la démarche

Pascale Montreer

► To cite this version:

Pascale Montreer. Identification des molécules responsables de l'odeur d'un mélange gazeux – Vers une automatisation de la démarche. Matériaux. IMT - MINES ALES - IMT - Mines Alès Ecole Mines - Télécom, 2019. Français. NNT : 2019EMAL0007 . tel-02899932

HAL Id: tel-02899932

<https://theses.hal.science/tel-02899932>

Submitted on 15 Jul 2020

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

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

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR D'IMT MINES ALÈS

En Chimie Informatique et statistique

École doctorale Risques et Société

Unité de recherche Odeurs et COV du Laboratoire du Génie de l'Environnement Industriel

IDENTIFICATION DES MOLECULES RESPONSABLES DE L'ODEUR D'UN MELANGE GAZEUX – Vers une automatisation de la démarche

Présentée par Pascale MONTREER

Le 5 novembre 2019

Sous la direction de Jean-Louis FANLO

Devant le jury composé de

El Mostafa QANNARI, Professeur, ONIRIS Nantes

Pierre LE CLOIREC, Professeur, Ecole Nationale Supérieure de Chimie de Rennes

Anne BERGERET, Professeur, IMT Mines Alès

Marie-France FALZON, Dr, Responsable du laboratoire d'analyses physico-chimiques,

HUTCHINSON, Chalette-sur-Loing

Mathilde CHAIGNAUD, Dr, Chargée d'études, OLENTICA, Alès

Jean-Louis FANLO, Professeur, IMT Mines Alès

Stefan JANAQI, Enseignant-chercheur, IMT Mines Alès

Stéphane CARIOU, Enseignant-chercheur, IMT Mines Alès

Rapporteur

Rapporteur

Présidente

Examineur

Examineur

Directeur de thèse

Co-Encadrant

Co-Encadrant

A Sarah,

*Qui nous a quitté beaucoup trop tôt et beaucoup trop soudainement
et qui, elle, n'aura jamais pu terminer sa thèse ...*

REMERCIEMENTS

Je tenais à remercier tout d'abord M. Jean-Louis Fanlo et Mme Isabelle Bétremieux sans qui ce projet de thèse n'aurait jamais vu le jour. Je les remercie grandement pour leur présence et leurs conseils pertinents pour ce projet et pour la suite. Je remercie également M. Emmanuel Gomez pour avoir également été à l'initiative du projet.

Je remercie M. Frédéric Picard, Mme Marie-France Falzon, M. Budagwa Assumani et Mme Sabine Sirol d'avoir encadré ce projet sur les trois années.

Je ne saurais oublier d'évoquer M. Philippe Ricoux qui a participé grandement à l'évolution du projet. Je le remercie pour ses conseils et sa culture statistique et mathématique.

Je remercie Sylvain Lozano et Maxime Durot pour leur temps et leur volonté de poursuivre le projet. Aussi, je remercie l'équipe Total DS de m'avoir accueilli dans leurs locaux régulièrement.

Je tenais également à remercier tout particulièrement M. El Mostafa Qannari et M. Pierre Le Cloirec d'avoir accepté d'être mes rapporteurs de thèse et ainsi d'avoir pris le temps de la lire intégralement. Je les remercie pour leur bienveillance.

Je tiens également à remercier Mme Anne Bergeret d'avoir accepté d'être présidente de mon jury de thèse.

Je remercie également Stéphane Cariou et Stefan Janaqi d'avoir encadré mon travail et partagé avec moi leurs compétences respectives.

Ensuite, je tenais à dire un grand merci à Mathilde Chaignaud pour sa grande implication dans toutes les mesures et manipulations qui ont été nécessaires à mon travail de thèse ainsi qu'à sa présence bienveillante sur l'ensemble du projet.

A présent, je tiens à dire un grand merci à toutes les personnes du laboratoire de l'IMT Mines Alès qui ont été présentes pour moi tout au long de ces cinq années et qui m'ont permis de tenir la distance. Je citerai tout particulièrement Cathy Balieu qui a été toujours présente et à mon écoute de manière très assidue, Sandrine Bayle pour son énergie, Perrine Branchet pour son calme et Emilie Marchand pour son optimisme à toutes épreuves ! Je remercie également Dédé, Sylvie, Roro, Janick et Mumu, pour leur bonne humeur et pour le café et les thés journaliers.

Je tiens à remercier tous les doctorants qui ont croisé mon chemin et que je n'oublierai jamais et notamment mes collaborateurs de l'ATHEMA : Alexandra Siot et son loto, Benjamin Chollet et son vin

chaud, Pierre-Baptiste Jacquot qui m'a tout appris d'un président ATHEMA, Brahim Mazian toujours à l'écoute. Je citerai également Clément Chanut, Alexandre Le Borgne, Thibault Jeannin, Cécile Lhéritier, Roland Eyssette et Quentin Perez, l'équipe de choc pour leur bonne humeur et leur humour (« Souvenirs souvenirs »). Sans oublier William Garat, Erwan Carré, Adrien Coutouis et Benoît Graillet,

Une pensée également pour Frédéric Heymes, Scherrine Tria, Romain Ravel, Benjamin Gallard, Marion Fages, Evelyne Touraud qui ont contribué aussi à ces bons moments pendant ma thèse.

Enfin je remercie mes proches qui ont su être là quand il le fallait.

RESUME

Dans le milieu industriel, les matériaux générant une odeur désagréable représentent une problématique majeure.

En effet, l'odeur fait très souvent partie des critères de sélection, d'achat et d'utilisation d'un produit par le consommateur. Si un matériau a une odeur désagréable, il risque d'être rejeté par le consommateur qui considérera sa qualité comme mauvaise ou altérée.

Améliorer la qualité odorante d'un matériau constitue donc un enjeu industriel et économique important.

Dans ce contexte, le travail réalisé dans le cadre de cette thèse consiste à développer un protocole permettant d'identifier le ou les composés chimiques responsables de l'odeur désagréable d'un matériau. Ce protocole est développé sous forme d'un outil automatisé combinant une succession de techniques statistiques.

L'un des piliers de ce travail est la recherche de corrélations entre la composition de la matrice gazeuse émise par le matériau (mesures physico-chimiques) et l'odeur associée à cette matrice (mesures olfactométriques). Pour atteindre cet objectif, un important travail d'investigation sur les seuils de perception olfactive (fiabilité et complétude des données) a été réalisé.

ABSTRACT

In the industrial environment, materials generating an unpleasant smell are a major problem.

Indeed, the smell is often part of the criteria for the selection, purchase and use of a product by the consumer. If a material has an unpleasant smell, it may be rejected by the consumer who will consider its quality as bad or distorted.

Improving the odorant quality of a material is therefore an important industrial and economic challenge.

In this context, the thesis project consists in developing an analytical protocol to identify the chemical compounds responsible for the unpleasant odor of a material. This analytical protocol is developed as an automated tool combining a succession of statistical techniques.

One of the pillars of this work is the search for correlations between the composition of the gaseous matrix emitted by the material (physico-chemical measurements) and the odour associated with this matrix (olfactometric measurements). To achieve this objective, an important investigation work on odour detection thresholds (reliability and completeness of data) was carried out.

TABLE DES MATIERES

REMERCIEMENTS	4
Résumé.....	7
Abstract	7
Table des matières.....	9
Abréviations	15
Introduction générale	17
CHAPITRE 1 : Identifier les molécules responsables de l'odeur d'un matériau : quelle démarche ?	19
1.1 Introduction	19
1.2 De l'olfaction à l'odeur.....	20
1.2.1 Le mécanisme de l'olfaction	20
1.2.2 Les dimensions de l'odeur.....	22
1.2.3 Évaluation des caractéristiques de l'odeur	26
1.2.4 Une notion fondamentale : le seuil de perception	28
1.3 De l'échantillonnage à l'identification et la quantification des molécules.....	30
1.3.1 L'échantillonnage.....	30
1.3.2 Les techniques d'analyse.....	33
1.4 L'approche de « l'expert » ou l'art du compromis.....	35
1.4.1 Conditionnement du matériau / échantillonnage	35
1.4.2 Caractérisations physicochimiques et sensorielles	36
1.4.3 Mise en relation des données physicochimiques et olfactométriques	38
1.5 Questions posées / Objectifs de l'étude	39
1.5.1 Variabilité liée à « l'expert »	40
1.5.2 Variabilité liée aux données utilisées	40
1.5.3 Objectifs de l'étude.....	41
CHAPITRE 2 : Fiabilisation de l'utilisation des données de seuils de perception olfactive 43	
2.1 Contexte.....	43
2.2 Démarche globale	45
2.3 Étape 1 : Constitution de la base de données.....	46
2.3.1 Définition de la base de données	46
2.3.2 Contenu	47
2.4 Étape 2 : Élimination des articles considérés comme les moins fiables.....	47
2.4.1 Méthodologie de l'étape 2.....	47
2.4.2 Résultats de l'étape 2.....	49
2.5 Étape 3 : Élimination de seuils de perception « aberrants » par molécule	49
2.5.1 Méthodologie de l'étape 3	49
2.5.2 Résultats de l'étape 3.....	54
2.6 Étape 4 : Élimination des articles « aberrants ».....	55

2.6.1	Méthodologie.....	55
2.6.2	Résultats de l'étape 4.....	56
2.7	Étape 5 : Évaluation de la qualité du nettoyage à partir d'un référentiel expérimental de seuils de perception.....	60
2.7.1	Méthodologie de l'étape 5.....	60
2.7.2	Résultats de l'étape 5.....	64
2.8	Conclusion sur le protocole de nettoyage.....	69

CHAPITRE 3 : Amélioration du degré de complétude des données de seuils de perception 73

3.1	Contexte.....	73
3.2	Choix de l'approche du modèle.....	74
3.2.1	Méthodologie.....	74
3.2.2	Les modèles de type RQSA/RQSP et leur mise en œuvre.....	74
3.3	Choix des indices.....	77
3.3.1	Critère de sélection des indices.....	77
3.3.2	Résultat : Base de données d'entrée.....	80
3.4	Sélection du jeu de construction.....	82
3.4.1	Méthodologie.....	82
3.4.2	Définition des jeux de construction et jeu de validation.....	86
3.5	Génération du modèle.....	98
3.5.1	Méthodologie de modélisation.....	98
3.5.2	Résultats de la modélisation.....	101
3.6	Affinage du modèle global par analyse des erreurs individuelles.....	105
3.6.1	Méthodologie d'analyse.....	105
3.6.2	Résultat de l'affinage du modèle global par analyse des erreurs.....	107
3.7	Évaluation et limites de la modélisation.....	118
3.7.1	Apport de l'utilisation des seuils de perception sans incertitude dans la conception du modèle.....	118
3.7.2	Limite d'application du modèle.....	121
3.8	Conclusion sur la complétude des seuils de perception manquants.....	125

CHAPITRE 4 : Mise au point d'une démarche automatisée d'identification des molécules responsables de l'odeur d'une matrice gazeuse..... 127

4.1	Contexte.....	127
4.2	Matériels et méthodes.....	128
4.2.1	Fabrication des mélanges gazeux synthétiques.....	128
4.2.2	Approche « expert » : Détermination des molécules responsables de l'odeur des mélanges.....	134
4.2.3	Démarche automatisée d'identification des molécules responsables de l'odeur.....	138
4.3	Résultats.....	142
4.3.1	Mise en pratique et paramétrage de la démarche automatisée.....	142
4.3.2	Impact de la dégradation du critère de similarité.....	145
4.4	Conclusion sur la démarche automatisée.....	147

CHAPITRE 5 : Impact de la précision des données utilisées sur l'identification des molécules responsables de l'odeur..... 149

5.1	Contexte.....	149
5.2	Évaluation de l'impact de la précision de la quantification.....	150
5.2.1	Méthodologie.....	150
5.2.2	Résultats.....	151

5.2.3	Bilan sur l'impact du mode de quantification	159
5.3	Évaluation de l'impact de la précision des valeurs de seuils de perception	160
5.3.1	Méthodologie.....	160
5.3.2	Résultats	162
5.3.3	Bilan sur le scénario d'obtention des seuils de perception	167
5.4	Évaluation de l'impact de la précision de l'analyse olfactométrique	167
5.4.1	Méthodologie.....	167
5.4.2	Résultats	168
5.5	Conclusion intermédiaire sur l'impact de la qualité des données utilisées sur la détermination des molécules responsables de l'odeur.....	172
CHAPITRE 6 : Exemple d'application de la démarche		175
6.1	Contexte général et objectif	175
6.2	Matériels et méthodes	175
6.2.1	Matériaux analysés	175
6.2.2	Protocole de l'étude.....	176
6.3	Résultats.....	178
6.3.1	Analyses physico-chimiques	178
6.3.1	Analyses olfactométriques.....	180
6.3.2	Seuils de perception.....	180
6.4	Discussion.....	182
6.5	Conclusion	183
CHAPITRE 7 : Conclusion générale, Discussion, et perspectives		185
7.1	Conclusion générale	185
7.2	Discussions	188
7.2.1	Les concepts	188
7.2.2	Les apports de la démarche	190
7.3	Limites et perspectives de la recherche	191
7.3.1	L'identification et la quantification des molécules.....	192
7.3.2	La gestion des données de seuils de perception existantes.....	192
7.3.3	La modélisation des seuils de perception	193
7.3.4	L'outil d'identification des molécules responsables de l'odeur	194
Table des figures		196
Table des tableaux.....		200
Bibliographie.....		204
Annexe 1	Résultats du calcul de la FE et du TC des articles.....	222
Annexe 2	Résultats de la classification des articles.....	223
Annexe 3	Seuils de perception expérimentaux : le référentiel	224
Annexe 4	Protocole de volatilisation d'un échantillon liquide.....	225
Annexe 5	HeatMap des valeurs obtenues suite au test de Dixon sur données du référentiel non standardisées.....	226

Annexe 6	Répétabilité des mesures expérimentales	227
Annexe 7	Ecarts absolus à la moyenne (EAM) individuels pour chaque molécule du référentiel après standardisation	228
Annexe 8	Valeurs de Z1, Z2, Re, Z1', Z2' et Re'	229
Annexe 9	Nombre d'observations pour chaque molécule de chaque jeu de données 230	
Annexe 10	Calcul des indices sélectionnés pour le modèle RSQA/RSQP	231
Annexe 11	Indices de structure des 350 molécules – Partie 1	235
Annexe 12	Indices de structure des 350 molécules – Partie 2	242
Annexe 13	Utilisation de l'indice d'Hopkins	249
Annexe 14	Coordonnées des molécules sur les axes PP (ordre selon Axe 1 croissant) 350 molécules.....	250
Annexe 15	Coordonnées des indices de structure sur les axes PP (ordre selon Axe 1 croissant) (47 indices).....	253
Annexe 16	Résultat de la classification des molécules à partir de leurs coordonnées sur les deux axes principaux de la « Projection Pursuit ».....	254
Annexe 17	Décomposition des classes selon les groupements fonctionnels (350 molécules)	259
Annexe 18	Modélisation globale du logarithme décimal du seuil de perception des 350 molécules	260
Annexe 19	Molécules supports parmi les 277 molécules du jeu de construction.	265
Annexe 20	Résultats totaux de la modélisation globale et affinée construite sur 350 molécules	266
Annexe 21	Données permettant d'évaluer la justesse et la précision de la fabrication d'un mélange liquide	271
Annexe 22	Exemple d'application du protocole de volatilisation d'un mélange liquide synthétique en mélange gazeux synthétique.....	272
Annexe 23	Composition de l'essai liquide A	274
Annexe 24	Analyse de répétabilité des analyses physico-chimiques.....	275
Annexe 25	Résultats PLS par mélange suivant le seuil de similitude utilisé.....	276

Annexe 26	Quantification des molécules du mélange F à partir des courants ioniques totaux par les différentes méthodes	279
Annexe 27	Quantification des molécules du mélange G à partir des courants ioniques totaux par les différentes méthodes	280
Annexe 28	Graphique des OAV des mélanges F et G en fonction de la quantification appliquée	281
Annexe 29	Graphes des VIP résultant des PLS appliquées aux études des mélanges F et G en fonction de la quantification appliquée	282
Annexe 30	Graphes des OAV des mélanges 2 et 5 en fonction des scénarios de seuils de perception appliqués	283
Annexe 31	Graphes des VIP résultant des PLS appliquées aux études des mélanges 2 et 5 en fonction des scénarios de seuils de perception appliqués	285
Annexe 32	Graphes des VIP résultant des PLS appliquées aux études des mélanges 2 et 5 en fonction du taux de variation de la concentration d'odeur expérimentale par rapport à la référence du jury 1	287
Annexe 33	Organigramme résumé de la démarche	288

ABREVIATIONS

I	: Intensité de l'odeur
CO	: Concentration d'Odeur
AO	: Activité d'Odeur
TIC	: Courant Ionique Total
QDA	: Qualitative Descriptive Analysis
MDS	: Multi Dimensional Scaling
BDD	: Base De Données
SP	: Seuil de Perception
EAM	: Ecart Absolu Moyen
PI	: Perte d'Information
TC	: Taux de comparaison
FE	: Fréquence d'Exclusion des auteurs
ODG	: Ordre de Grandeur
QSAR	: Quantitative Structure-Activity Relationship (Relation quantitative structure-activité)
QSPR	: Quantitative Structure-Property Relationship (Relation quantitative propriété-activité)
InChI	: International Chemical Identifier
SMILES	: Simplified Molecular Input Line Entry Specification
CEP	: weighted electronic connectivity matrix
CAH	: Classification Ascendante Hiérarchique
PP	: Projection Pursuit (ou Poursuite de Projection)
ACI	: Analyse en Composante Indépendante
CV	: Coefficient de Variation
ER	: Erreur relative
PLS	: « Partial Least Square regression » ou régression des moindres carrés partiels
VIP	: « Variable Importance Projection » ou variable d'importance dans la construction de la variable réponse

INTRODUCTION GENERALE

Nous percevons notre environnement avec nos cinq sens : la vue (Réception d'énergie électromagnétique), l'ouïe et le toucher (Mécano-réception), ainsi que le goût et l'odorat (Chimio-réception). Ces cinq sens permettent à l'être humain de répondre de manière adéquate aux divers stimuli de son environnement. Ces stimuli peuvent, par exemple, provoquer des sensations d'appétit, de joie, de danger, ou encore de dégoût. Nos perceptions sont croisées (CANDAU et JEANJEAN 2006). Les informations apportées par les différents sens peuvent être complémentaires, redondantes, voire contradictoires.

Quel que soit le contexte, les sens alertés en permanence sont la vue, l'odorat et l'ouïe. En effet, ces trois sens captent en permanence les signaux de l'environnement de manière « spontanée ». Plus indirectement, le toucher transmet des informations sur la température ou l'humidité.

Paradoxalement, si l'odorat est généralement considéré comme un sens mineur (LLEDO et VINCENT 1999), il est aussi reconnu comme le plus subtil de nos cinq sens. Les nuisances odorantes sont, avec les nuisances sonores, les plus immédiatement ressenties et les plus mal tolérées par la population. La perception d'une odeur à caractère désagréable par un individu est en effet toujours assimilée à une dégradation de la qualité de son environnement, à une mauvaise qualité de l'objet qui en est à l'origine, voire à un risque pour sa santé. À ce titre, les nuisances odorantes constituent le deuxième motif de plaintes dans la population après celles liées au bruit (SAULNIER et ZANIN 2003) et l'on constate aujourd'hui à une sensibilisation croissante des populations (MONTESTRUCQ et al. 2007).

Cette problématique est par voie de conséquence devenue très structurante pour le monde industriel en général, et pour celui de la fabrication des matériaux en particulier. L'odeur fait en effet très souvent partie des critères de sélection, d'achat et d'utilisation d'un produit. Il arrive fréquemment qu'une odeur désagréable soit à l'origine d'un rejet par l'utilisateur final.

La recherche de solutions pour diminuer l'impact des rejets gazeux odorants sur les populations riveraines des sites industriels, ou pour améliorer l'odeur d'un matériau, nécessite de connaître les caractéristiques des odeurs et de comprendre quelles molécules peuvent expliquer le caractère désagréable de l'odeur. Sur la base de cette connaissance, des actions préventives ou correctives peuvent alors être imaginées et mises en œuvre pour améliorer le confort des riverains ou la perception de la qualité du produit par les consommateurs.

La pollution odorante a fait l'objet de nombreuses études, et des méthodes d'évaluation et de gestion des odeurs ont été mises en place (ATMOTERRA 2014). Définir une action ou un traitement pour

limiter l'impact de l'odeur reste toutefois plus difficile à mettre en place que pour le cas d'une pollution sonore. L'odeur est en effet une perception plus complexe à appréhender que le son.

Les sources de pollution odorantes sont nombreuses et diverses. Il peut s'agir de sites industriels aussi variés que des sites chimiques, de compostage, d'équarrissage, des élevages mais également de matériaux utilisés dans la vie quotidienne tels que revêtements de sols ou muraux, matières plastiques présentes dans les habitacles de moyens de transport, etc.

Dans ce contexte, le travail présenté dans ce manuscrit a pour objectif de développer un protocole analytique permettant d'identifier le ou les composés chimiques responsables de l'odeur désagréable d'un matériau. Ce protocole analytique sera développé sous forme d'un outil automatisé combinant une succession de techniques statistiques.

L'un des piliers de ce travail est la mise en relation de la composition de la matrice gazeuse émise par le matériau (caractérisée par des mesures physico-chimiques) et l'odeur associée (caractérisée par des mesures olfactométriques).

Le chapitre qui suit est consacré à l'exposé de la problématique et à la description de l'approche actuellement utilisée pour rechercher l'origine de l'odeur de matériaux (approche « expert »). Il explique dans un premier temps ce qu'est l'odeur, quelles sont ses caractéristiques et les méthodes de mesure associées avant de présenter la démarche de l'expert, ses différentes alternatives et les questions sous-tendues.

CHAPITRE 1 : IDENTIFIER LES MOLECULES RESPONSABLES DE L'ODEUR D'UN MATERIAU : QUELLE DEMARCHE ?

1.1 Introduction

Comme mentionné plus haut, les odeurs influencent nos choix car elles jouent un rôle de premier plan dans l'appréciation de la qualité, du confort et du bien-être.

Le mécanisme par lequel un matériau présente une odeur vient du fait que certaines molécules contenues dans la matrice solide ont une pression de vapeur saturante suffisamment élevée pour être présentes en concentrations significatives dans la phase gazeuse en équilibre avec le solide. Parmi ces molécules, certaines, dites « odorantes », sont susceptibles de déclencher la perception d'une odeur (agréable ou non) par un observateur.

Le premier défi consiste donc à identifier dans la matrice gazeuse généralement complexe la ou les molécules pouvant expliquer la force de l'odeur, sa persistance, sa qualité ou son caractère éventuellement désagréable. Relever ce premier défi nécessite de réunir des compétences en analyse physicochimique (il faut identifier et quantifier les molécules présentes le plus précisément et le plus exhaustivement possible), mais aussi en analyse sensorielle (il faut caractériser l'odeur par toutes ses dimensions). L'objectif étant à ce stade d'identifier les molécules responsables de l'odeur, il est souhaitable que ces compétences soient présentes chez l'expert en charge de l'analyse.

Le deuxième défi est de déterminer l'origine de ces molécules. S'agit-il de molécules constitutives de la matrice solide, de monomères résiduels, d'impuretés, de produits néoformés ? La compétence des polyméristes, des formulateurs et des transformateurs est ici essentielle pour orienter la réflexion.

Enfin, le troisième défi, et non le moindre, consiste à imaginer des actions correctives et à vérifier leur efficacité. On le voit, cette boucle, idéalement vertueuse, n'est pas loin de ressembler à un cercle vicieux, tant les problèmes posés sont complexes et les réponses peu évidentes.

Apporter une réponse complète au problème de l'odeur désagréable d'un matériau implique de relever les trois défis qui viennent d'être cités.

Répondre à la question posée dans le titre de ce chapitre présuppose de relever le premier défi mentionné. S'il ne s'agit là que de la première étape du processus global, cette étape est pourtant fondamentale, car la qualité des réponses apportées à ce stade conditionnera celle des étapes suivantes.

Pour bien appréhender les enjeux et problèmes liés à l'identification des molécules responsables de l'odeur dans une matrice gazeuse, il faut avoir une vision claire de ce qu'est une odeur, de la façon dont on la perçoit et de ses caractéristiques mesurables. La première partie de ce chapitre est donc consacrée à ce sujet.

La section qui suit est destinée à la description du fonctionnement du système olfactif, des caractéristiques de l'odeur, et à un état de l'art sur les méthodes permettant d'évaluer ces caractéristiques. Puis, une section est consacrée à la description de l'analyse d'une matrice gazeuse. La démarche actuellement utilisée par les experts dans ce domaine est ensuite exposée. Enfin, l'objectif global de la thèse est présenté.

1.2 De l'olfaction à l'odeur

1.2.1 Le mécanisme de l'olfaction

Notre système olfactif permet à la fois la circulation de l'air jusqu'aux poumons (la ventilation) et la détection des molécules odorantes présentes dans l'air que nous respirons. Cette détection est assurée grâce au contact entre ces molécules et le dispositif de détection constitué par la muqueuse olfactive située au sommet et au fond des fosses nasales dans la cavité olfactive (Figure 1).

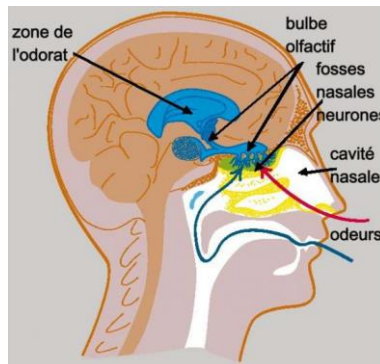


Figure 1 : Anatomie du système olfactif (GANTY 2016)

Nous inspirons en moyenne 0,5 litre d'air quinze à seize fois par minute, soit 10 à 12 m^3 d'air par jour. Les molécules odorantes peuvent atteindre la muqueuse olfactive par deux voies : la voie directe (inspiration par le nez), et la voie rétro-nasale (expiration par le nez entraînant vers la muqueuse l'air contenu dans la bouche). Lors d'une inspiration normale, 10 à 15 % de cet air atteint la muqueuse olfactive.

La muqueuse olfactive (Figure 2) est constituée de trois types de cellules (MORRISON et COSTANZO 1990):

- Les cellules olfactives qui sont des neurones bipolaires. L'une de leurs extrémités (la dendrite) se dirige vers la cavité nasale et porte un bouquet de cils chimio sensibles qui baignent dans la couche de mucus sécrété par les glandes de Bowman. L'autre extrémité (l'axone) se dirige vers le bulbe olfactif. La membrane de ces cils comporte des récepteurs moléculaires qui peuvent se lier aux molécules odorantes et créer un signal vers le cerveau.
- Les cellules de soutien qui facilitent le développement des neurones, mais ne jouent pas de rôle dans la transduction du signal olfactif.

- Les cellules basales qui sont des cellules souches qui peuvent se différencier soit en cellules olfactives, soit en cellules de soutien.

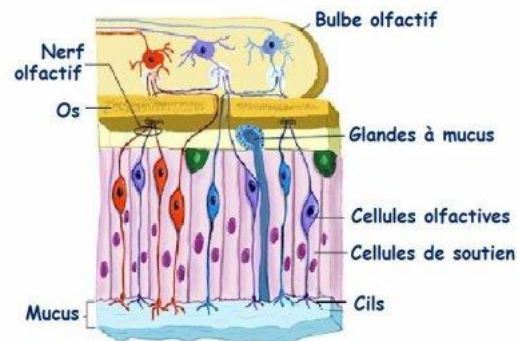


Figure 2 : Structure de la muqueuse olfactive

La muqueuse olfactive a chez l'homme une surface moyenne de l'ordre de 2 à 3 cm^2 avec environ 10000 cellules olfactives par mm^2 .

Le mucus sécrété par les glandes de Bowman a pour fonction de protéger les cellules réceptrices, mais aussi d'assurer la captation et le transport des molécules odorantes.

L'interaction des molécules odorantes avec les protéines réceptrices situées sur les cils est très courte (quelques millièmes de seconde). Le stimulus chimique est transformé en un signal électrique dans la cellule sensorielle (BUCK et AXEL 1991). C'est l'étape de transduction. Le signal ainsi créé se propage alors le long de l'axone, puis de la dendrite, vers le cerveau (DE MARCH 2015).

La dernière étape est l'étape de codage de l'information, qui correspond au traitement du signal électrique dans le bulbe olfactif et dans le cerveau. Dans le bulbe olfactif, le signal est transformé en une « image » que le cerveau traite comme une image visuelle. Le mécanisme par lequel le cerveau traite ce signal est caractérisé par un codage à deux niveaux : l'odeur est identifiée en faisant appel aux zones de la cognition et de la mémoire et l'intensité est appréciée via la fréquence des impulsions électriques. Le mécanisme global de l'olfaction est résumé en Figure 3.

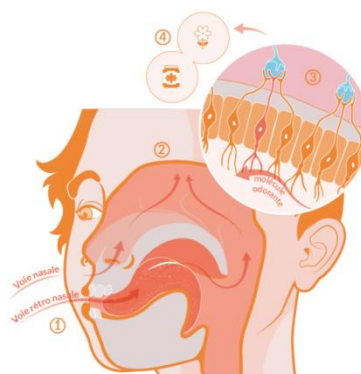


Figure 3 : Mécanisme de l'olfaction (GANTY 2016)

Le mécanisme par lequel une molécule ou un mélange gazeux génère une odeur correspond donc à une succession de processus chimiques, physiologiques, mnésiques et cognitifs (ROGNON & POURTIER, 2014).

En conclusion, l'odeur est une perception, fruit de l'interprétation par le cerveau d'un signal électrique issu de l'interaction entre un mélange de composés volatils et la muqueuse olfactive. Elle ne peut en aucun cas être considérée comme une propriété intrinsèque de l'objet odorant.

1.2.2 Les dimensions de l'odeur

La perception olfactive (l'odeur), qu'elle soit liée à la présence d'une seule molécule dans l'air ou à un mélange complexe, peut être caractérisée par quatre grandeurs (DELVA, COBUT, et FANLO 2017):

- Son acceptabilité
- Sa qualité
- Son intensité
- Sa persistance

Ces quatre grandeurs sont présentées et définies dans les parties qui suivent.

1.2.2.1 L'acceptabilité/ton hédonique

Le ton hédonique représente la tonalité affective qui mesure le caractère plaisant ou déplaisant de l'odeur : son acceptabilité.

Une odeur peut être agréable, acceptable, désagréable, intolérable. Les échelles d'acceptabilité peuvent être discrètes ou continues [(VDI-3882-2 1994) ; (AFNOR 2012) ; (CHAIGNAUD et al. 2014)].

Ce classement est subjectif, l'accoutumance pouvant être à l'origine d'un glissement de l'appréciation de l'observateur vers la classe acceptable et le jugement porté dépendant, pour une substance donnée, de la concentration de l'odeur. Généralement, il existe un assez bon accord, dans une population homogène, sur les odeurs considérées comme agréables, désagréables ou intolérables, bien que l'on constate, là encore, des différences d'appréciation individuelles, souvent liées à des habitudes et à des cultures différentes [(E.P. KOSTER 1991) ; (SUCKER et al. 2008)].

1.2.2.2 La qualité de l'odeur

Déterminer la qualité de l'odeur revient à déterminer sa nature, ses évocations et ce à quoi elle s'apparente (ROGNON & POURTIER, 2014). Cette étape permet de décrire l'odeur à l'aide de descripteurs sensoriels afin d'établir son profil sensoriel. La qualité de l'odeur est la première caractéristique de l'odeur identifiée par le cerveau, même si le premier réflexe de l'individu qui perçoit l'odeur est de fournir une information sur l'acceptabilité (« ça sent bon/mauvais ») plutôt que l'identification [(BENOIT et PANNER 1982) ; (MAILLE 2000)].

Un signal odorant peut être perçu très différemment d'un individu à l'autre sur le plan qualitatif, en fonction notamment du passé propre de chacun ou encore de leur degré de connaissance des odeurs. La qualité reste à ce titre la propriété de l'odeur la plus subjective car elle dépend du vécu (variation inter individuelle) [(ROUBY et SICARD 1997) ; (DAVID et al. 1997)]. La qualification des odeurs souffre aussi d'une déficience du vocabulaire à cause du manque de mots pour définir l'odeur (ROUBY et al. 2005). Ainsi, si un champ sémantique n'est pas fixé lors de la description qualitative d'une odeur, il sera possible de se trouver confronté à des évocations comme : « odeur de Lego® » ; « odeur de piscine neuve » ; ou encore « plastique » et « vanille » pour décrire une même odeur. Un travail de consensus sera alors nécessaire pour signifier que ces évocations sont effectivement associées à une odeur de plastique et de sucre.

1.2.2.3 L'intensité

L'intensité d'une odeur correspond à la force de la perception [(ROGNON et POURTIER 2000) ; (LESTREMAU 2003) ; (NICOLAS, ROMAIN, et DELVA 2008) ; (SUCKER et al. 2008) ; (PIERRETTE et MOCH 2009) ; (NICELL 2009)]. La relation qui permet d'établir le lien entre l'intensité du stimulus (la concentration du mélange odorant) et l'intensité de la perception olfactive est la loi de Stevens (Équation 1). C'est une loi psychophysique analogue à d'autres régissant des modalités sensorielles comme la vision ou l'audition (WRIGHT 1978).

$$I = K \cdot C^n$$

I = intensité perçue
C = concentration dans l'air inhalé (ppm)
K = constante

Équation 1 : Loi de Stevens

Une fois linéarisée, l'Équation 1 devient l'Équation 2.

$$\log I = n \log (C - C_0) + Cte$$

I = intensité perçue
C = concentration dans l'air inhalé (ppm)
C₀ = concentration au seuil de perception : c'est la concentration donnant une odeur juste décelable

Équation 2 : Loi de Stevens linéarisée

Le coefficient "n" (généralement compris entre 0,2 et 0,8) traduit l'accroissement de l'intensité odorante en fonction de la concentration.

Chaque odeur peut donc se caractériser par une courbe que l'on peut tracer entre l'intensité de l'odeur et sa concentration (courbe log-log). Cette courbe présente trois parties (Figure 4):

1. Un « palier d'inodorité » qui correspond à la zone dans laquelle la concentration des molécules dans l'air inhalé ne permet pas de détecter l'odeur (intensité nulle) ;

2. Une branche montante (de pente « n ») qui traduit l'accroissement de l'intensité de la perception avec l'augmentation de la concentration ou encore la pente de persistance ;
3. Un palier de saturation qui correspond à la zone au-delà de laquelle l'intensité de la perception ne croît plus avec l'augmentation de la concentration.

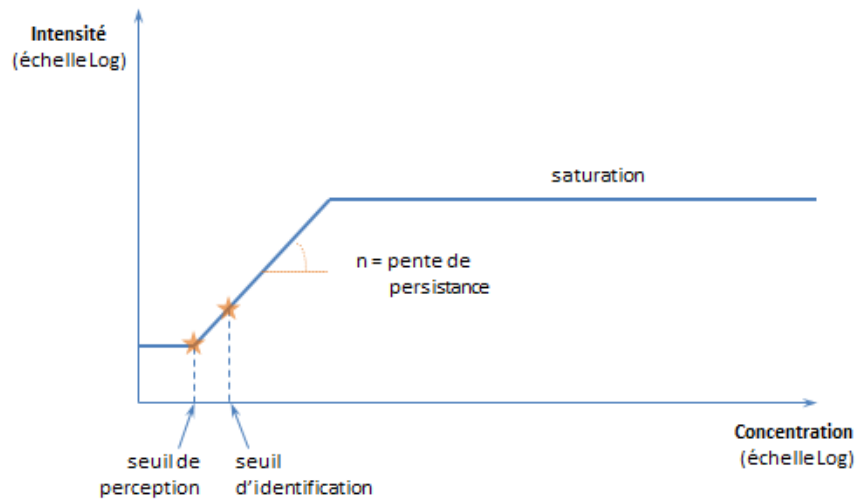


Figure 4 : Courbe de Stevens

Il est important de noter que les caractéristiques de l'odeur qui viennent d'être présentées ne sont pas indépendantes : la qualité d'une odeur, et par voie de conséquence son acceptabilité, peuvent changer selon son intensité. L'odeur du jasmin, par exemple, dont le caractère floral et subtil domine à faible concentration se modifie à forte concentration avec l'apparition du caractère animal et désagréable. Généralement, une augmentation de l'intensité d'une odeur s'accompagne d'une diminution de son acceptabilité (DISTEL et al. 1999). Il existe toutefois des exceptions : la vanilline est agréable indépendamment de la concentration à laquelle elle est présentée.

1.2.2.4 La persistance

La persistance représente le lien entre la concentration des molécules odorantes et l'intensité de la perception. C'est la pente de la branche montante de la courbe de Stevens. La persistance correspond donc à la sensibilité de l'odeur à la dilution (Figure 5).

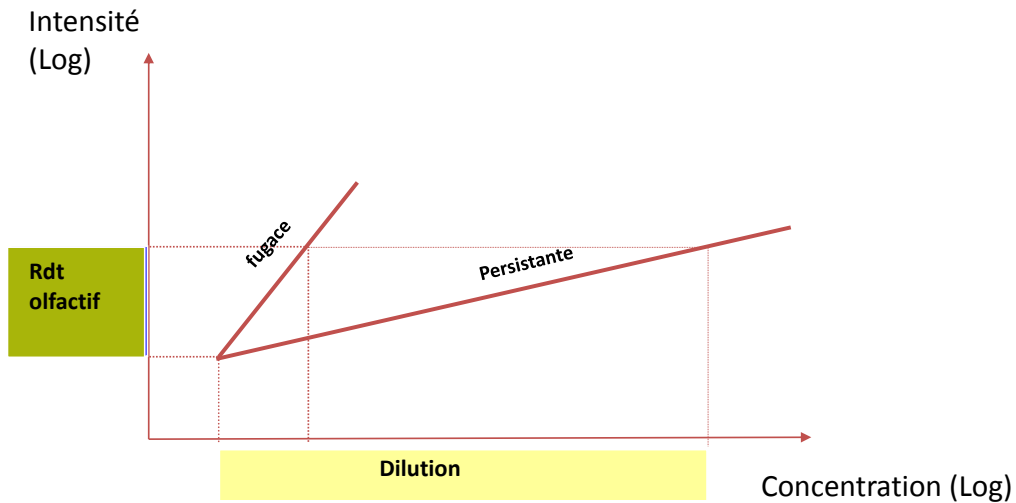


Figure 5 : Odeur persistante/ odeur fugace

Si n est fort (0,8), l'odeur est dite fugace car son intensité diminuera rapidement lorsque la concentration des molécules dans l'air diminuera. Si n est faible (0,2), l'odeur sera dite persistante car son intensité restera quasi constante pour la même diminution de concentration (Figure 5).

La persistance est évaluée par la mesure de la concentration d'odeur qui correspond au facteur de dilution nécessaire pour atteindre le seuil de perception (AFNOR 2003). Plus précisément, on recherche, par dilution de l'air odorant avec de l'air inodore, le taux de dilution nécessaire pour que 50% des membres d'un jury perçoivent l'odeur, c'est-à-dire le seuil de perception. Lorsqu'il est atteint, le nombre de dilutions est alors directement converti en concentration d'odeur, exprimée en Unités d'Odeur Européenne par mètre-cube (UO_E/m^3).

Cette notion est importante dans la recherche de solution de traitement des odeurs. La Figure 6, permet de l'illustrer.

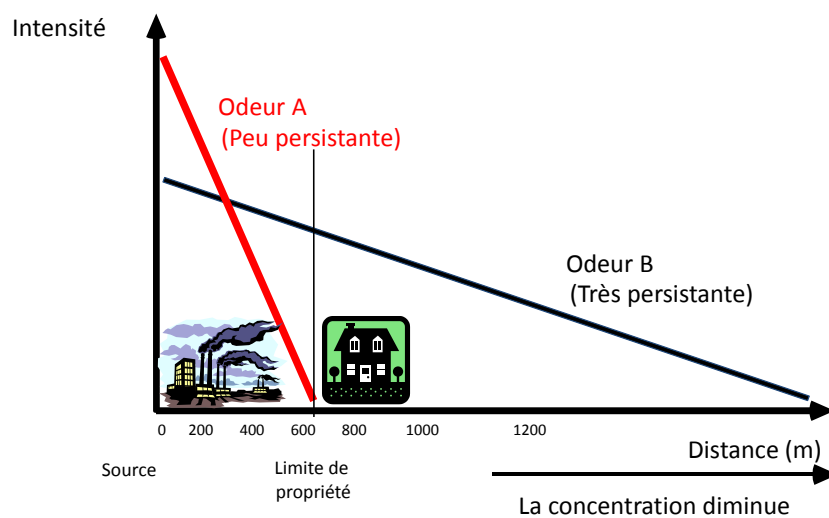


Figure 6 : Évolution de l'intensité d'une odeur en fonction de la distance à la source émettant l'odeur

Ainsi, dans cet exemple, même si, sur le site, l'odeur A apparaît plus intense que l'odeur B, il est préférable de traiter l'odeur B pour améliorer le confort olfactif des riverains, cette dernière présentant une caractéristique persistante.

1.2.3 Évaluation des caractéristiques de l'odeur

Les méthodes d'évaluation des dimensions présentées précédemment sont décrites ci-après.

1.2.3.1 L'acceptabilité/ton hédonique

Il n'existe pas d'échelle universellement acceptée pour évaluer le caractère agréable ou désagréable d'une odeur. Le guide européen VDI 3882-2 :1994 propose une des échelles les plus utilisées, avec une classification de 9 points (Tableau 1). Cette classification a été reprise par plusieurs auteurs comme Van Harreveld (VAN HARREVELD 2001), l'Environmental Protection Agency d'Irlande (EPA - Environmental Protection Agency 2001) et Sucker et al. (SUCKER et al. 2008).

Acceptabilité/Ton hédonique - VDI 3882-2	
+4 : extrêmement agréable	-4 : extrêmement désagréable
+3 : très agréable	-3 : très désagréable
+2 : agréable	-2 : désagréable
+1 : peu agréable	-1 : peu désagréable
0 : neutre	

Tableau 1 : Classification du ton hédonique (VDI-3882-2 1994)

Olentica (CHAIGNAUD et al. 2014) a récemment développé une méthode d'évaluation de l'acceptabilité de l'odeur qui présente l'intérêt de permettre, une fois la mesure de la concentration de l'odeur réalisée, de caractériser l'acceptabilité à l'aide du même échantillon et avec le même jury. La méthode présente l'avantage de ne pas soumettre les jurés à des concentrations d'odeur importantes et de fournir un niveau d'acceptabilité pour l'échantillon d'essai sur une échelle ouverte.

1.2.3.2 La qualité de l'odeur

Le but de cette évaluation est d'obtenir le profil olfactif de chaque odeur. Deux approches sont possibles pour obtenir cette information, l'une orientée, l'autre libre (ou non orientée).

Dans l'approche orientée, les panelistes sont formés à décrire les odeurs à l'aide d'un champ lexical contraint. Il s'agit d'approches de type roue des odeurs [(SUFFET et ROSENFELD 2007) ; (DECOTTIGNIES et al. 2013)] ou champ des odeurs® [(JAUBERT, GORDON, et DORE 1987a) ; (JAUBERT, GORDON, et DORE 1987b)].

Dans l'approche non orientée il n'est pas nécessaire d'apprendre un champ lexical. La description de la qualité de l'odeur se fait en langage naturel. Une méthode de ce type est développée par Medjkoune

(MEDJKOUNE et al. 2016). Elle est basée sur le traitement du langage naturel et sur les techniques de représentation des connaissances. Elle peut donc être utilisée pour évaluer les odeurs avec toute personne intéressée.

1.2.3.3 L'intensité

La mesure de l'intensité odorante est décrite dans la norme française NF X 43-103 (AFNOR 1996). La méthode consiste à comparer l'intensité de l'odeur d'une atmosphère donnée à celles d'une échelle de référence composée de dilutions de produits de référence (pyridine ou n-butanol). L'échelle de dilution est constituée sur la base d'une progression géométrique d'un facteur 2, 3 ou 10. La mesure est réalisée grâce à un jury de nez, idéalement composé de 7 à 9 personnes.

L'objectif de chaque membre du jury est d'estimer à quelle dilution de la gamme correspond la force de l'odeur sentie. Une conversion est ensuite réalisée en équivalent de l'échelle de concentration de la référence utilisée.

1.2.3.4 La persistance

La persistance de l'odeur est évaluée par la mesure de ce qui est appelé dans la norme européenne EN 13725 (AFNOR 2003) la « concentration de l'odeur ». La concentration de l'odeur est définie comme étant le nombre de fois qu'il faut diluer le mélange odorant avec de l'air pur pour arriver au seuil de détection d'une personne présentant une sensibilité olfactive considérée comme normale (au sens statistique), donc représentative de la population moyenne. La norme impose que les jurés doivent percevoir l'odeur d'une molécule de référence, le n-butanol, à une concentration comprise entre 20 et 80 ppb. Cette perception doit en outre être constante dans le temps.

L'unité de la concentration de l'odeur est l'unité d'odeur européenne par mètre cube (UO_E/m^3). Selon la norme, l'unité d'odeur européenne est « la quantité de substance odorante qui, évaporée dans un mètre cube de gaz neutre, dans des conditions normalisées, déclenche une réponse physiologique de la part d'un jury (seuil de détection) équivalente à celle suscitée par une masse d'odeur de référence européenne, évaporée dans un mètre cube de gaz neutre dans des conditions normalisées ».

Comme mentionné plus haut, la référence est à ce jour basée sur le n-butanol, molécule dont le seuil de perception est bien renseigné dans la littérature. Une unité d'odeur européenne par mètre cube (UO_E/m^3) correspond à une concentration de $123\mu\text{g}/\text{m}^3$ de n-butanol.

La mesure de la concentration d'odeur est réalisée à l'aide d'un olfactomètre à dilution dynamique (Figure 7 et Figure 8).

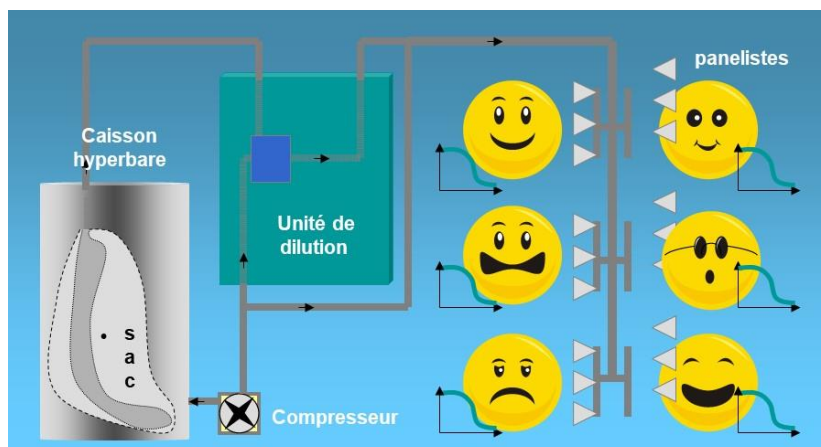


Figure 7 : Schéma de principe d'un olfactomètre à dilution dynamique (Doc Olentica®)



Figure 8 : Olfactomètre à dilution dynamique (ODILE®)

Cet instrument permet de diluer le mélange odorant de façon dynamique et de présenter la dilution à un jury de nez. Les membres du jury doivent parvenir à faire la distinction entre de l'air inodore et une dilution de l'odeur étudiée qui leur est présentée. Au départ, le mélange odorant leur est présenté à un niveau de dilution tel qu'ils n'arriveront pas à faire la distinction avec l'air inodore. Le taux de dilution est ensuite progressivement diminué. Cette opération est répétée jusqu'à atteindre le seuil de perception du juré.

1.2.4 Une notion fondamentale : le seuil de perception

Il existe, pour chaque molécule, un seuil de concentration à partir duquel un individu est capable de faire la différence entre de l'air pur et de l'air contenant cette concentration de la molécule considérée. C'est le seuil de perception olfactive. Le Tableau 2 fournit quelques exemples de seuils de perception pour des familles de molécules bien connues pour leurs propriétés odorantes.

Composés	Formule	Seuil de perception (mg/m ³)
Composés soufrés		
Hydrogène sulfuré	H ₂ S	6,95 10 ⁻⁰⁴
Sulfure de carbone	CS ₂	2,39 10 ⁻⁰²
Diméthyl sulfure	CH ₃ -S-CH ₃	2,54 10 ⁻⁰³
Composés azotés		
Ammoniac	NH ₃	2,66 10 ⁻⁰²
Diméthylamine	(CH ₃) ₂ NH	8,47 10 ⁻⁰²
Triméthylamine	(CH ₃) ₃ N	1,11 10 ⁻⁰⁴
Indole	C ₈ H ₆ NH	2,87 10 ⁻⁰³
Scatole	C ₉ H ₈ NH	7,50 10 ⁻⁰⁴
Cadavérine	NH ₂ (CH ₂) ₅ NH ₂	7,51 10 ⁻⁰⁴
Acides gras volatils		
Acide acétique	CH ₃ COOH	2,50 10 ⁺⁰⁰
Acide butyrique	CH ₃ (CH ₂) ₂ COOH	1,08 10 ⁻⁰³
Acide valérique	CH ₃ (CH ₂) ₃ COOH	2,50 10 ⁻⁰³
Aldéhydes		
Acétaldéhyde	CH ₃ CHO	1,80 10 ⁻⁰⁴
Formaldéhyde	HCHO	1,23 10 ⁻⁰¹
Butyraldéhyde	C ₃ H ₇ CHO	3,83 10 ⁻⁰²
Propionaldéhyde	CH ₃ CH ₂ CHO	1,71 10 ⁻⁰¹

Tableau 2 : Exemples de seuils de perception d'après « Pollutions Olfactives » de l'ADEME (ADEME 2005)

À ce niveau de concentration, l'individu ne perçoit pas la qualité de l'odeur (a fortiori son éventuel caractère agréable ou désagréable). La réponse qu'il peut fournir est donc de type binaire (je perçois une différence avec de l'air pur ou non), ce qui autorise une exploitation statistique des résultats.

Dès lors que la concentration de la molécule dans l'air dépasse ce seuil, la qualité de l'odeur, et par voie de conséquence son acceptabilité, deviennent perceptibles.

Le seuil de perception apparait donc comme une donnée d'importance cruciale dans la détermination des molécules susceptibles de jouer un rôle dans l'odeur : une molécule dont la concentration dans un mélange gazeux est supérieure ou égale à son seuil de perception olfactive sera susceptible de participer à l'odeur globale du mélange. La notion d' « activité d'odeur » (Équation 3) traduit ce principe (ROTHER et THOMAS 1963).

$$AO_i = \frac{C_i}{SP_i}$$

AO_i : Activité d'odeur de la molécule i (adimensionnel)
C_i : Concentration massique du composé i (mg/m³)
SP_i : Seuil de perception olfactive de la molécule i (mg/m³)

Équation 3 : Définition de l'activité d'odeur d'une molécule i

Pour illustrer l'importance de cette grandeur, on peut prendre l'exemple d'un mélange binaire gazeux constitué d'acide acétique et d'acide valérique à concentrations égales (5 mg/m³). Le premier possède un seuil de perception de 2,5 mg/m³ alors que celui du second est de 0,0025 mg/m³.

L'activité d'odeur de l'acide valérique est de 2000 ($\frac{5}{0,0025} = 2000$) alors que celle de l'acide acétique est de 2 ($\frac{5}{2,5} = 2$). Cela signifie qu'il y a de très fortes chances pour que l'acide valérique soit seul responsable de l'odeur du mélange.

1.3 De l'échantillonnage à l'identification et la quantification des molécules

Comme mentionné au début de ce chapitre, la démarche qui vise à identifier les molécules à l'origine de l'odeur d'un matériau nécessite un double niveau d'investigation puisqu'il faut mettre en relation les différentes grandeurs caractéristiques de l'odeur avec la composition physico-chimique du gaz. Cela passe dans un premier temps par la mise en œuvre de techniques d'échantillonnage et d'analyse plus ou moins sophistiquées. Les concentrations de molécules volatiles peuvent être mesurées sur la base de prélèvements ponctuels simples ou accumulés dans le temps (préconcentrations) suivis d'une analyse ou bien par une surveillance directe continue en temps réel. L'analyse des molécules volatiles passe ainsi par deux étapes obligatoires. L'échantillonnage est la première. La seconde consiste en l'analyse proprement dite (qualitative et/ou quantitative).

Les techniques d'échantillonnage et d'analyse les plus couramment utilisées dans l'approche des composés volatils émis par les matériaux sont présentées ci-après.

1.3.1 L'échantillonnage

Il existe deux types de techniques d'échantillonnage :

- les prélèvements sans concentration
- les techniques de concentration

Le type d'échantillonnage adopté dépend généralement de la méthode analytique qui sera ensuite employée et des concentrations requises pour obtenir des résultats (TUDURI 2002).

1.3.1.1 Prélèvements sans concentration

Cette approche consiste à introduire l'échantillon dans un contenant pour son transport et sa conservation jusqu'à l'analyse. L'objectif principal est de maintenir l'intégrité de l'échantillon, c'est-à-dire d'éviter les réactions, les pertes par fuites ou adsorption sur les parois et d'éviter les contaminations par le matériel de stockage utilisé.

Plusieurs dispositifs sont envisageables : ampoules en verre, conteneurs métalliques, sacs en matériaux plastiques.

1.3.1.1.1 Ampoules en verre

Les ampoules en verre sont de volumes variables (0,05 à 2 L) équipées de 2 robinets d'arrêts en Téflon ou en verre et d'un orifice pouvant disposer d'un septum (Figure 9).

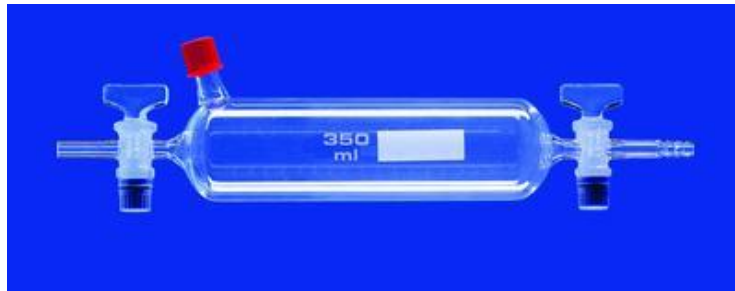


Figure 9 : Ampoule de prélèvement

L'échantillonnage est réalisé en connectant une des extrémités à une pompe qui aspire l'air à travers l'ampoule. Une fois l'air initialement présent dans l'ampoule remplacé par l'air à analyser, les robinets sont fermés. L'air peut ensuite être prélevé à l'aide d'une seringue à travers le septum pour être directement injecté en chromatographie phase gazeuse.

Ce type de dispositif permet de préserver l'intégrité de l'échantillon mais le faible volume des ampoules limite son utilisation en couplage avec les techniques de pré-concentration.

1.3.1.1.2 Les conteneurs métalliques (Canisters)

Il s'agit de conteneurs en acier inoxydable dans lesquels l'échantillon est soit aspiré directement par mise sous vide préalable du conteneur, soit aspiré via une pompe. L'air peut dans ce dernier cas être comprimé ce qui permet d'échantillonner un grand volume (Figure 10). Il est possible en prélevant à faibles débits, d'obtenir des données moyennes de concentrations dans le temps. Ce type de conteneur permet d'obtenir une grande stabilité des concentrations de composés volatils (jusqu'à trente jours après leur prélèvement). Leur principal défaut est d'être encombrants, lourds et chers (KOGA et al. 2001).



Figure 10 : Canister

1.3.1.1.3 Sacs d'échantillonnage

Les sacs d'échantillonnage sont les plus couramment utilisés. Leur utilisation nécessite de prendre en compte les phénomènes de diffusion à travers les parois et d'adsorption sur les parois du sac qui dictent le choix des matériaux et la durée de conservation des échantillons. Ils sont à base de polymères tels que le Tedlar® (fluorure de polyvinyle), le Téflon® (polytétrafluoroéthylène), le Mylar® (type particulier de polyester) ou le Nalophan® (polyéthylène téréphtalate) (Figure 11).



Figure 11 : Sacs d'échantillonnage

Le principal avantage de cette solution est qu'elle permet la collecte d'un grand volume d'échantillon qui peut être analysé ultérieurement par plusieurs techniques. Cet aspect est important, car il permet de réduire les sources de variabilité dans le processus global d'analyse.

Toutes ces techniques de prélèvement sans concentration peuvent être suivies par une analyse directe ou par une étape intermédiaire de préconcentration avant analyse. Si elles sont mises en œuvre sans couplage à une technique de concentration de l'échantillon, les composés volatils de la source d'émission doivent être à une concentration suffisamment élevée pour être détectés par la technique d'analyse, sinon le facteur de concentration possible sera dépendant du volume prélevé. Dans le contexte particulier des émissions de matériaux, les concentrations sont généralement faibles, voire très faibles et l'échantillonnage sera la plupart du temps suivi par un processus de concentration (NICOLLE 2009).

1.3.1.2 Les techniques de concentration

Le but de la mise en œuvre des techniques de concentration est d'augmenter la concentration des molécules présentes dans l'échantillon en modifiant le moins possible leurs concentrations relatives pour se placer au-dessus de la limite de détection des appareils d'analyse utilisés.

1.3.1.2.1 L'espace de tête (ou headspace)

La technique de l'espace de tête est couramment utilisée pour le prélèvement de composés volatils à partir d'un échantillon liquide ou solide positionné dans un récipient clos jusqu'à ce que les composés volatils parviennent à un équilibre entre l'échantillon et la phase gazeuse constituant l'espace de tête. Un volume déterminé de cet espace de tête est ensuite prélevé pour l'analyse (TRANCHANT 1996).

Il existe deux types d'espace de tête : l'espace de tête statique et l'espace de tête dynamique ou « purge-and-trap » (FERNANDEZ, BREME, et VARLET 2009).

Le premier consiste à prélever la phase gazeuse qui se trouve en équilibre thermodynamique dans un système clos (TRANCHANT 1996). Cette méthode est simple mais la sensibilité qui en résulte est limitée. Son utilisation est donc limitée à des milieux riches concentrés (D'ACAMPORA ZELLNER et al. 2008).

La technique de l'espace de tête dynamique consiste à libérer la totalité des composés volatils présents dans un échantillon liquide en retirant de façon continue la phase gazeuse du système (TRANCHANT

1996). Cette méthode permet d'augmenter la sensibilité de l'analyse [(GRIMM, CHAMPAGNE, et OHTSUBO 2001) ; (BOUTOU et CHATONNET 2007) ; (MUNOZ et al. 2010)].

Généralement l'espace de tête est re-concentré avant l'analyse par d'autres techniques de concentration. La limitation principale de cette technique dans le contexte qui nous concerne est la taille limitée des échantillons qu'elle permet d'analyser, souvent mal adaptée à la problématique des matériaux.

Dans le domaine de l'analyse des molécules volatiles émises par des matériaux, les micro-chambres constituent un outil intéressant (NICOLLE 2009). Ces micro-chambres permettent de porter le matériau à des températures comprises entre l'ambiance et 250 °C. Elles sont balayées en continu par un débit d'air ou de gaz vecteur contrôlé. Les composés émis par les matériaux sont recueillis sur un tube d'adsorbant puis analysés en TD/GC/MS.

1.3.1.2.2 Les adsorbants

Les adsorbants sont des matériaux poreux à la surface desquels les molécules d'intérêt peuvent être piégées par des interactions de faible énergie. Il existe une grande diversité de matériaux adsorbants qui sont utilisés seuls ou en combinaison (DESAUZIERS 2009). Chaque adsorbant possède, en fonction de ses caractéristiques chimiques et de celles de sa porosité, une affinité préférentielle pour certaines familles de molécules. Le choix de l'adsorbant dépend donc de la nature (polarité, taille, volatilité) des molécules à concentrer.

Les molécules piégées subiront ensuite une désorption thermique qui permettra de les transférer vers l'analyseur (DESAUZIERS 2009).

1.3.1.2.3 La micro extraction en phase solide (SPME)

La microextraction sur phase solide (SPME) est une technique d'adsorption/désorption permettant de concentrer les composés volatils d'échantillons liquides ou gazeux. L'adsorbant est à la surface d'une fibre de silice constituant l'extrémité de l'aiguille d'une seringue. La fibre est directement thermodésorbée dans l'injecteur du chromatographe. La SPME est soumise aux mêmes contraintes que l'utilisation des adsorbants en tubes [(TUDURI 2002) ; (LESTREMAU 2003)].

1.3.2 Les techniques d'analyse

Ce paragraphe n'a pas l'objectif de fournir une présentation exhaustive des techniques d'analyses de composés gazeux. Il se borne à évoquer ci-dessous uniquement les techniques d'analyse les plus habituellement utilisées dans le cadre des problématiques d'odeur.

1.3.2.1 Les couplages GC-MS et GCXGC-MS

La technique la plus utilisée est la chromatographie en phase gazeuse couplée à la spectrométrie de masse [(DAVOLI et al. 2003) ; (BOUTOU et CHATONNET 2007) ; (RODRIGUEZ-NAVAS, FORTEZA, et CERDA 2012) ; (GALLEGO et al. 2012)]. Elle permet la séparation des constituants de

mélanges gazeux complexes par une suite continue d'équilibres entre une phase mobile gazeuse et une phase stationnaire et l'identification des molécules séparées.

Au plan qualitatif, la GC-MS permet d'obtenir des identifications reproductibles, identifiable pour des quantités injectées de l'ordre de 10^{-10} g par molécules. Au plan qualitatif, ses limites de détection sont parmi les plus basses, et des quantifications précises sont possibles à condition de pouvoir réaliser un étalonnage sur chaque molécule à quantifier.

La séparation des composés constitutifs du mélange peut être améliorée en recourant à la chromatographie gazeuse dite bidimensionnelle. La séparation des molécules volatiles se fait dans ce cas à travers deux colonnes capillaires, placées l'une à la suite de l'autre dans le(s) four(s) d'un appareil de chromatographie, et éventuellement couplées par un système de modulation.

Si la totalité de l'éluât provenant de la colonne de séparation primaire subit une nouvelle séparation, on parle de GC 2D « comprehensive » (CORDERO et al. 2015). Si ce sont seulement des fractions bien déterminées de l'éluât qui subissent la deuxième séparation, on parle de « heart cutting » [(LO et al. 2008) ; (PARAVISINI et al. 2014)].

1.3.2.2 Le couplage chromatographie en phase gazeuse/spectrométrie de masse/olfactométrie (GC/MS/O)

Dans le couplage GC/MS/O, le nez humain est utilisé comme détecteur sensoriel, en complément de la détection par spectrométrie de masse décrite ci-dessus. Ce couplage permet de détecter et d'identifier les molécules responsables d'une odeur (D'ACAMPORA ZELLNER et al. 2008). L'odeur étant une perception humaine, l'analyse olfactive est réalisée par un panel d'opérateurs : chaque opérateur sent en sortie de colonne, matérialise le moment où il perçoit une odeur, évalue l'intensité de la perception et en fournit une description de la qualité.

1.3.2.3 Les nez électroniques

Les nez électroniques sont des dispositifs multi-capteurs qui permettent de détecter des molécules dans l'air. Il est important de souligner ici que ces dispositifs mesurent la concentration des molécules volatiles présentes dans l'air et non l'odeur, qui est une perception et non une propriété intrinsèque de l'objet odorant (§1.2.1). La réponse de l'ensemble des capteurs utilisés dans le nez électronique est corrélée avec des mesures olfactométriques [(MILAN, BOOTSMA, et BILSEN 2012) ; (GUTIERREZ et al. 2014) ; (MILAN, BOOTSMA, et BILSEN 2016)].

Ces systèmes sont surtout utilisés lorsque les concentrations doivent être surveillées en continu (surveillance environnementale, contrôle de procédé de production).

1.3.2.4 La spectrométrie de masse à résonance protonique (PTR-MS)

La spectrométrie de masse à résonance protonique est une technique de spectrométrie de masse qui ouvre la possibilité de suivre en temps réel plusieurs dizaines de molécules dans l'air à des niveaux de

concentration ppb-ppt (partie par billion-par trillion) (VESIN et al. 2012). Cette caractéristique originale permet de s'affranchir des techniques de préconcentration (tubes d'adsorbants/SPME) qui génèrent des biais souvent incontournables dans l'analyse des échantillons gazeux (spécificité des adsorbants) et ne permettent pas d'avoir accès à la dynamique d'évolution des concentrations car elles fournissent des valeurs de concentrations moyennes sur la durée du prélèvement (plusieurs minutes à plusieurs dizaines de minutes). L'accès à la dynamique d'évolution des concentrations est d'un grand intérêt dans l'étude des phénomènes transitoires (odeurs, émissions de matériaux, temps de réponse de capteurs, ...) [(LINDINGER et al. 2008) ; (K. H. HAN et al. 2011) ; (SCHUHFRIED et al. 2012)].

Il s'agit d'une technique de spectrométrie de masse basée sur une ionisation douce des molécules contenues dans des échantillons gazeux. L'énergie d'ionisation est modulable en fonction du gaz d'ionisation utilisé (H_2O , NO, O_2 , Xe, Kr). Dans tous les cas, il n'y a pas (ou peu) de fragmentation des molécules, ce qui autorise un suivi en continu et une quantification directe.

Les avantages principaux de cette technique sont qu'elle fournit une analyse quantitative en temps réel (temps de réponse < 100 ms), la limite de détection en temps réel la plus basse du marché, et qu'elle ne nécessite aucune préparation d'échantillon (échantillonnage direct de l'air).

1.4 L'approche de « l'expert » ou l'art du compromis

Le problème de l'expertise nécessaire pour aborder le problème de la génération d'une odeur à caractère désagréable par un matériau a été abordé dans l'introduction de ce chapitre. Le niveau de cette expertise dépend, comme cela a été évoqué, de l'étendue du problème posé. Sur un plan général, un expert est une personne maîtrisant la connaissance dans le ou les domaines utiles à la réalisation de l'expertise et possédant une expérience reconnue qui lui permet d'évaluer les éléments faisant partie de son champ d'expertise. Dans le contexte qui est celui qui nous concerne, il se doit, au minimum, de maîtriser la métrologie des gaz et celle des odeurs. Confronté au challenge de l'identification des molécules à l'origine de l'odeur d'un matériau, l'expert doit en premier lieu choisir parmi les techniques d'échantillonnage et d'analyse celles qui satisfont le mieux aux contraintes d'une approche globale, à la fois physicochimique et sensorielle. Dans le cadre de ce travail, l'expert était une personne de la société Olentica. Ce sont les méthodes utilisées par cet expert et les raisons qui l'ont conduit à ces choix qui sont développées ci-après.

1.4.1 Conditionnement du matériau / échantillonnage

Pour atteindre son objectif et faire le lien entre la composition d'une phase gazeuse et l'odeur associée, l'expert est dans un premier temps confronté à des problèmes très pratiques. Il a besoin d'une analyse physicochimique la plus précise possible, en termes d'exhaustivité de l'identification des molécules présentes comme de quantification, mais aussi d'analyses olfactométriques elles aussi fiables.

Le premier problème posé par ce double niveau d'exigence réside dans la taille de l'échantillon gazeux nécessaire. En effet, si les différentes techniques d'analyse physico-chimiques à sa disposition (§1.3.2) ne nécessitent pas un échantillon de grand volume, il n'en va pas de même des analyses olfactométriques. Ces dernières rendent nécessaire de disposer d'un volume d'échantillon gazeux conséquent (quelques dizaines de litres) pour évaluer les différentes caractéristiques de l'odeur (§1.2.2). La résolution de ce premier problème passe par le choix d'une technique d'échantillonnage adaptée. À la contrainte qui vient d'être évoquée se rajoute le fait que la minimisation des sources de variabilité dans les résultats des différentes analyses passe par la possibilité de les réaliser sur le même échantillon. Cela explique donc que la technique de conditionnement du matériau la plus utilisée par l'expert soit celle des sacs de prélèvement. Elle seule en effet permet de constituer un échantillon gazeux de taille suffisante pour faire face aux différentes analyses devant être réalisées. Cette technique est généralement mise en œuvre dans un local à température régulée.

La constitution de l'échantillon gazeux se fait en introduisant une masse, un volume ou une surface connue du matériau à étudier dans un sac Nalophan® dans lequel est ajouté un volume connu d'air inodore. Le sac est positionné dans une pièce à température régulée jusqu'à l'atteinte de l'équilibre thermodynamique.

L'atteinte de cet équilibre est vérifiée par des mesures régulières de la concentration en COV à l'aide d'un analyseur à photoionisation (PPB RAE Systems) (Figure 12).



Figure 12 : Détecteur à photoionisation (PPB RAE System)

Une fois l'équilibre atteint, le gaz présent dans le sac est soumis aux analyses.

1.4.2 Caractérisations physicochimiques et sensorielles

Les analyses sensorielles permettent de déterminer trois caractéristiques de l'odeur : La persistance (concentration d'odeur selon la norme EN 13725), l'acceptabilité et la qualité (méthode Olentica). Les principes de ces méthodes ont été exposés au paragraphe 1.2.3.

Ces méthodes sont des méthodes psychophysiques qui permettent d'étudier les relations quantitatives entre les stimulations physiques et les sensations, ou plus généralement les réponses, qu'elles engendrent. Elles sont entachées par des niveaux d'incertitude élevés liés au fait qu'elles utilisent l'être humain comme instrument de mesure en mettant à profit ses capacités olfactives, en l'occurrence, pour

caractériser et évaluer des produits, notamment dans les domaines où les sens apportent une certaine valeur ajoutée par rapport aux mesures physico-chimiques usuelles.

La caractérisation physico-chimique de la matrice gazeuse est réalisée par une analyse GC-MS (§ 1.3.2.1) sur le même échantillon que celui utilisé pour la caractérisation de l'odeur, après concentration sur un tube d'adsorbants. Le tube constitué est désorbé thermiquement à l'aide d'un thermodésorbiteur et les molécules désorbées reconcentrées sur un piège froid maintenu à -30°C (cryofocalisation). À l'issue de cette désorption primaire, le piège froid subit un flash thermique pour permettre l'injection dans le chromatographe couplé à un spectromètre de masse.

Le choix de la technique de préconcentration sur tubes d'adsorbants et de celle d'analyse (GC-MS) est le résultat d'un premier compromis, dicté par des considérations pratiques : Elle est en effet l'une des techniques les plus éprouvées à ce jour, et elle permet d'atteindre des niveaux de sensibilité compatibles avec la détection des molécules odorantes. Elle présente pourtant plusieurs biais et défauts :

- la nécessité de préconcentrer les molécules par adsorption conduit inévitablement à une déformation de l'empreinte de la matrice gazeuse du fait de la spécificité des adsorbants utilisés ;
- la technique de séparation (chromatographie en phase gazeuse) ne permet pas toujours la séparation de tous les constituants du mélange gazeux (co-élutions possibles) ;
- la quantification est délicate à réaliser, sauf à disposer de standards de toutes les molécules identifiées.

À l'issue de cette étape, l'expert dispose donc de deux blocs de données :

- Un bloc de données issu des analyses olfactométriques : ce bloc comporte un faible nombre de données (une valeur de concentration d'odeur, une d'acceptabilité et quelques descripteurs de la qualité) caractérisées, comme mentionné plus haut, par un niveau d'incertitude élevé.
- Un bloc de données physicochimiques comportant des données beaucoup plus nombreuses (nature et concentration des différentes molécules) qui, même si elles sont elles aussi parfois biaisées par les caractéristiques des techniques utilisées pour les obtenir, sont plus précises et dont l'incertitude est mieux maîtrisée que celle des données psycho-sensorielles.

L'étape suivante consiste à mettre en relation ces deux blocs de données pour en tirer le maximum d'informations utiles.

1.4.3 Mise en relation des données physicochimiques et olfactométriques

Comme nous l'avons vu plus haut, l'expert va croiser les données physicochimiques et les données olfactométriques en se servant d'un troisième bloc de données : les valeurs de seuils de perception olfactive des molécules (§1.2.4). Ces informations clés pour son analyse sont recherchées dans la littérature [(LEONARDOS, KENDALL, et BARNARD 1969) ; (RUTH 1986) ; (DEVOS et al. 1990) ; (US EPA 1992) ; (ASTM 1997) ; (VAN GEMERT 2011)].

Il se trouve alors confronté à plusieurs cas de figure :

- Une valeur unique de seuil de perception de la molécule recherchée est disponible ;
- Plusieurs valeurs de seuil sont renseignées pour cette molécule ;
- Le seuil de perception n'est pas renseigné.

Ces trois cas de figure renvoient à des questionnements de différents niveaux :

Dans le premier cas, la question qui se pose est celle de la pertinence de cette valeur. La littérature citée en référence fournit en effet, comme cela sera développé dans le chapitre 2, des valeurs issues d'auteurs, de techniques, et d'époques très différentes.

Dans le deuxième cas, deux cas de figure peuvent se présenter :

- Les valeurs renseignées pour cette molécule sont toutes du même ordre de grandeur. L'expert calcule alors la moyenne géométrique de ces valeurs et considère cette moyenne comme représentative du seuil de perception moyen de la molécule.
- Les valeurs renseignées pour cette molécule s'étalent sur plusieurs ordres de grandeur (voir cas de l'acétate de butyle au paragraphe 2.1). L'expert est alors confronté à la nécessité de faire le choix parmi ces valeurs de celles qui lui paraissent les plus réalistes pour la molécule considérée. Il s'appuie pour faire ce choix sur sa propre expertise et sur des comparaisons qu'il effectue avec les seuils de perception de molécules présentant le plus de similarité structurale possible avec la molécule dont le seuil est recherché. Sur la base de cette analyse, il élimine alors les seuils qu'il juge « aberrants » et calcule la moyenne géométrique des seuils restants.

Dans le troisième cas, lorsque le seuil de perception est absent de la littérature, l'expert n'a d'autre choix que d'attribuer à cette molécule un seuil de perception en s'appuyant sur l'hypothèse que le seuil de perception de la molécule doit se situer dans la fourchette de ceux de la famille chimique à laquelle cette molécule appartient. Il identifie donc la famille chimique à laquelle appartient cette molécule, recueille les seuils de perception de ses membres, en fait la moyenne géométrique, et attribue – faute de mieux - cette valeur à sa molécule.

Une des limites de cette approche est qu'affilier une molécule à une famille chimique donnée n'est pas toujours chose facile en particulier lorsque cette molécule est polyfonctionnelle. En effet, si certains ouvrages de référence ou sites classent les molécules odorantes par familles chimiques, il s'agit le plus souvent de familles chimiques simples (alcane, alcène, aldéhyde, cétone, éther, ...).

Prenons pour illustrer cette difficulté l'exemple d'un expert identifiant la présence de méthoxyacétaldéhyde, molécule présentant une fonction aldéhyde et une fonction éther (Figure 13), dans le mélange gazeux qu'il analyse. Aucune information de seuil de perception n'est disponible pour cette molécule.

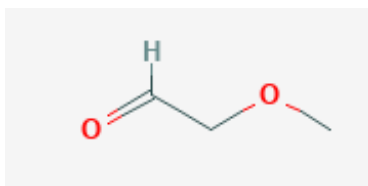


Figure 13 : Structure du méthoxyacétaldéhyde

L'expert a plusieurs possibilités : Procéder, comme précédemment, par analogie structurale en recherchant des molécules de structure comparable dont les seuils sont renseignés, attribuer à cette molécule le seuil de perception moyen de la famille des aldéhydes, ou des éthers, voire de l'ignorer.

À l'issue de sa démarche, l'expert a pu sélectionner un ensemble de molécules potentiellement responsables de l'odeur sur la base du calcul de leurs activités d'odeur (§1.2.4), c'est-à-dire sur la base de leur capacité à expliquer la concentration d'odeur du mélange gazeux (§1.2.2.4).

On le voit donc, la démarche que l'expert met en œuvre pour identifier les molécules à l'origine d'une odeur est complexe, nécessitant des compétences variées, en analyses physicochimiques et sensorielles, mais aussi en traitement de données de nature par essence même très différentes. Si elle permet d'identifier un certain nombre de molécules pertinentes, les résultats obtenus peuvent être très variables, car elle présente plusieurs inconvénients liés tant aux données manipulées qu'à l'expert lui-même.

Les problèmes liés à cette démarche et les objectifs de ce travail sont présentés ci-après.

1.5 Questions posées / Objectifs de l'étude

Comme nous venons de le voir, l'approche de l'expert passe par plusieurs niveaux de concessions faites face aux contraintes générées par les différents blocs de données manipulés, physicochimiques, olfactométriques et bibliographiques (les seuils de perception). Ces concessions soulèvent un certain nombre de questionnements qui seront détaillés ci-dessous.

La qualité du résultat de l'analyse de l'expert dépend donc de celle des données dont il dispose, mais avant tout de la manière dont il les utilise. Il est donc important d'examiner en quoi l'expert lui-même peut être source de variabilité des résultats.

1.5.1 Variabilité liée à « l'expert »

La première source de variabilité des résultats liée à l'expert réside à la base dans la disparité des compétences des experts eux-mêmes. La notion « d'expert » est en effet particulièrement floue, et par là même galvaudée. Dans le contexte qui est le nôtre, l'expert doit posséder des connaissances solides dans différents domaines. Dans celui de l'analyse physicochimique, afin de lui permettre d'avoir une conscience nette des biais et limites des techniques qu'il utilise. Même chose dans le domaine des analyses sensorielles. De plus, et surtout, il doit posséder une solide connaissance de la nature des matrices gazeuses qu'il analyse et des molécules odorantes qui peuvent y être présentes. Autant dire qu'il est rare de retrouver ces compétences réunies chez un même individu.

Toutes compétences égales par ailleurs, la qualité du résultat de l'analyse de l'expert dépend aussi de la finesse de son analyse, de sa « granularité ». Les molécules à l'origine de l'odeur peuvent en effet être les molécules majoritaires en termes de concentration, cas rare, mais beaucoup plus fréquemment des molécules minoritaires, présentes à l'état de traces dans la matrice analysée. En admettant que l'expert pousse son analyse jusqu'à considérer ces molécules, quelle sera sa démarche pour faire face à la variabilité et à l'incertitude liées à l'ensemble des données ? En effet, certaines, comme les seuils de perception, sont rarement disponibles, et lorsqu'elles le sont, ont un niveau de fiabilité aléatoire.

Il est clair que cette démarche, si elle se veut être conduite de façon rigoureuse, est particulièrement exigeante et chronophage : le nombre de molécules à traiter est tellement important qu'il se confronte aux biais et aux limites de détection des techniques habituellement utilisées, à l'incertitude et à l'incomplétude des bases de données sensorielles, et au temps que l'expert est en mesure de consacrer à l'analyse considérée.

Au-delà de ces considérations, il apparaît clairement que l'industriel porteur du problème est à ce jour dépendant d'une expertise dont le niveau est très difficile à évaluer, voire à maîtriser.

Ces considérations expliquent pourquoi nous avons choisi d'aborder cette problématique par une nouvelle approche, statistique et automatisée. L'enjeu de cette approche est de pallier au mieux les inconvénients qui viennent d'être cités.

1.5.2 Variabilité liée aux données utilisées

Les données utilisées sont, comme évoqué précédemment, elles aussi sources de variabilité. Les deux premiers blocs de données que croise l'expert sont les concentrations des molécules identifiées (données physicochimiques) et les seuils de perception de ces molécules (données bibliographiques). Il divise la concentration de chaque molécule par son seuil de perception pour calculer son activité d'odeur (§1.2.4). Il doit ensuite chercher les correspondances entre les molécules identifiées sur cette base comme responsables de l'odeur et les mesures olfactométriques réalisées qui constituent le troisième bloc de données. Il est donc clair que la fiabilité de ces données est primordiale. Les analyses physicochimiques

sont soumises à des biais et incertitudes inhérents aux principes de préconcentration, de séparation et de quantification des molécules. Les données bibliographiques, lorsqu'elles sont disponibles, sont souvent très incertaines. Les analyses olfactométriques qui permettent d'évaluer les différentes caractéristiques de l'odeur font partie du domaine des analyses psycho-physiques et sont à ce titre caractérisées par une incertitude beaucoup plus élevée que celle des analyses physicochimiques.

La manipulation de ces trois jeux de données, qui présentent des tailles, des incertitudes et des niveaux de complétude très différents pose une question fondamentale :

- Quel est le degré de précision nécessaire et suffisant sur ces trois blocs de données pour aboutir à une identification satisfaisante des molécules responsables de l'odeur ?

Cette question principale en sous-tend de nombreuses autres à l'échelle de chacun des blocs de données :

- Jusqu'à quel point est-il nécessaire de pousser l'identification des composées du mélange ?
- Quel est le niveau de précision nécessaire sur la quantification ?
- Quel est le niveau de précision nécessaire sur les analyses olfactométriques ?
- Comment, quand elles sont disponibles, juger de la fiabilité des données bibliographiques ?
- Comment gérer l'incomplétude de ces données ?

1.5.3 Objectifs de l'étude

Dans ce contexte, l'objectif de cette étude est de développer un système expert en remplaçant l'interprétation de l'expert par un système de traitement de données automatisé basé sur une boîte à outils de traitements mathématiques.

Cette approche statistique a pour ambition principale de réduire les sources d'incertitude et les inconvénients liés à l'approche déterministe de l'expert. En effet, si l'incertitude liée aux trois blocs de données nécessaires à l'analyse devra être prise en compte, la qualité de l'identification des molécules responsables de l'odeur ne dépendra plus que de la pertinence de la boîte à outils statistiques utilisée, et non de la compétence ou de la granularité de l'analyse de l'expert.

Cette approche permettra donc un gain substantiel de temps (méthode automatisée), une uniformité de traitement des résultats et offrira la possibilité d'appropriation de la méthode par l'industriel concerné par la problématique.

La suite de la présentation de ce travail comporte quatre chapitres.

Le chapitre 2 traite de la fiabilisation des données de seuils de perception disponibles dans la littérature. La complétude de ces mêmes données est abordée dans le chapitre 3. La mise en place de l'outil permettant d'identifier les molécules responsables de l'odeur est décrite en chapitre 4. Le chapitre 5 traite du problème de l'influence du degré de précision des analyses physico-chimiques et olfactométriques sur le résultat de l'identification des molécules responsables de l'odeur. Le chapitre 6 conclut ensuite par une application de la méthodologie mise au point à un cas concret (réel). Un schéma présentant la démarche est présenté en Annexe 33.

CHAPITRE 2 : FIABILISATION DE L'UTILISATION DES DONNEES DE SEUILS DE PERCEPTION OLFACTIVE

2.1 Contexte

Comme cela a été expliqué dans le chapitre précédent, le seuil de perception est un élément clé de l'identification des molécules responsables de l'odeur d'un mélange gazeux odorant.

On trouve dans la littérature au moins une valeur de seuil de perception olfactive pour seulement 650 molécules [(LEONARDOS, KENDALL, et BARNARD 1969) ; (RUTH 1986) ; (DEVOS et al. 1990) ; (US EPA 1992) ; (ASTM 1997) ; (VAN GEMERT 2011)].

Un nombre très important de molécules n'est donc tout simplement pas renseigné. Lorsque cette grandeur a été mesurée pour une molécule, deux cas de figure se présentent. Soit la molécule est renseignée un faible nombre de fois et la mesure présente par conséquent une fiabilité faible, soit la molécule est renseignée un nombre important de fois mais les seuils de perception présentent le plus souvent une forte variabilité. Cette variabilité peut être illustrée par le cas de l'acétate de butyle (butyl acetate) (VAN GEMERT 2011) présenté dans le Tableau 3.

Tableau 3 : Variabilité des valeurs de seuils de perception de l'acétate de butyle (VAN GEMERT 2011)

Références	Valeurs de seuils de perception de l'acétate de butyle (mg/m^3)
(JUNG 1936)	0,044
(MAY 1966)	35,000
(Egon Peter KOSTER 1971)	480,000
(HELLMAN et SMALL 1974)	0,030
(DRAVNIKS 1974)	3,000
(ANON 1980)	0,320
(NAUS 1982)	0,700
(NAGY 1991)	1,000
(PATTERSON et al. 1993)	7,700
(COMETTO-MUNIZ et CAIN 1994)	11,500
(ZIEMER, WOO, et ANAGNOSTOU 2000)	0,061
(NAGATA 2003)	0,077
(COMETTO-MUNIZ, CAIN, et ABRAHAM 2004)	0,015
(CAIN et SCHMIDT 2009)	0,010

Les valeurs de seuils de perception sont ici comprises dans un intervalle allant de $0,030 mg/m^3$ à $480 mg/m^3$ sur 14 articles, soit un ratio de 10^4 entre le maximum et le minimum.

L'importante variabilité des données de seuil de perception peut s'expliquer par plusieurs facteurs.

Tout d'abord, quel que soit le pays, aucune norme ne définit l'intégralité du protocole de mesure d'un seuil de perception. Seule est définie la mesure de la concentration d'odeur. Cela constitue la première source de variabilité.

La deuxième source de variabilité peut être liée aux zones géographiques dans lesquelles sont réalisées les mesures. En effet, les articles sont majoritairement issus du Japon, de Corée, d'Allemagne, d'Angleterre, du Danemark, des Pays-Bas et de France. Or, les protocoles de mesure de la concentration

d'odeur sont régis par des réglementations différentes dans ces pays. Au Japon, par exemple, la mesure de la concentration d'odeur est régie par la loi sur le contrôle des odeurs [(HIGUCHI 2004) ; (UENO et al. 2009)], en Corée par la loi sur la conservation de l'air (S. B. YANG 2003) et en Europe par la norme EN 13725 [(AFNOR 2003) ; (SNEATH 2003)]. Chaque réglementation présente ses propres spécificités et implique que les auteurs européens et nord-américains utilisent l'olfactométrie dynamique (choix forcé ou mode Oui / Non) alors que les auteurs japonais utilisent, eux, la « Triangle odor bag method » (UENO et al. 2009). Une comparaison de ces méthodes a été menée par différents auteurs [(AMANO et UENO 2004) ; (UENO et al. 2004) ; (UENO et al. 2009)]. Il a été observé que les valeurs de seuil de perception obtenues par olfactométrie dynamique en mode choix forcé ont tendance à être plus élevées (de 0,5 sur l'échelle logarithmique de concentration) que celles obtenues par la méthode japonaise. Cependant, les valeurs sont identiques entre la méthode japonaise et l'olfactométrie dynamique en mode Oui / Non. L'autre information apportée par Ueno est que lorsque le panel de jurés est sélectionné selon le critère de la norme EN 13725 pour les deux méthodes, les valeurs de seuils de perception obtenues par la méthode japonaise sont identiques à celles obtenues par olfactométrie dynamique à choix forcé (UENO et al. 2009).

En effet, la normalisation européenne impose comme critère de sélection des membres du panel un test de sensibilité olfactive au n-butanol (butan-1-ol), alors que la méthode japonaise est, elle, basée sur cinq molécules différentes: alcool β -phényléthylique (2-phenylethanol), méthyl cyclopenténolone (2-hydroxy-3-methylcyclopent-2-en-1-one), acide isovalérique (3-methylbutanoic acid), γ -undécalactone (5-heptyloxolan-2-one) et scatole (3-methyl-1H-indole) (SAIKI 2003). Comme cela est mentionné par Ueno (UENO et al., 2009), cette différence de méthodologie est source de variabilité.

La variabilité observée pour les valeurs de seuils de perception olfactive peut aussi être due à la nature statique ou dynamique de la méthode, ou au mode de présentation des dilutions (progressif ou aléatoire) (FALCY et MALARD 2005).

La définition même du seuil peut être source d'incertitude et de variabilité. En effet, si la majorité des auteurs considèrent que le seuil de perception est atteint lorsque 50% des membres du panel ont détecté l'odeur (norme EN 13725, (AFNOR 2003)), d'autres considèrent la valeur de 100% de détection du panel (LEONARDOS, KENDALL, et BARNARD 1969).

La qualité de l'échantillon sur lequel est pratiquée la mesure est une autre source de variabilité. Par qualité, on entend ici la pureté de la molécule utilisée et la précision avec laquelle est constitué l'échantillon gazeux qui va servir de base à la mesure (FALCY et MALARD 2005).

Le panel lui-même constitue également une source potentielle de variabilité. En effet, même si chacun de ses membres satisfait aux exigences de la norme, la qualité de la mesure dépend entre autres de l'implication des jurés (DRAVNIIEKS et PROKOP 1975), de leur sensibilité individuelle à la molécule étudiée (FALCY et MALARD 2005), de leur connaissance ou non de la qualité d'odeur de la molécule

étudiée (LEONARDOS, KENDALL, et BARNARD 1969) et de leur niveau de fatigue. Il est possible, et prévu dans certaines normes, de gérer cette variabilité. Le tri rétrospectif des réponses dans la normalisation européenne EN13725 (AFNOR 2003) ou la correction par l'intensité des odeurs dans la normalisation japonaise (UENO et al. 2009) en sont des exemples. Toutefois, ces traitements complémentaires ne sont pas systématiquement appliqués.

Il faut souligner ici que les valeurs que nous avons collecté correspondent à des études réalisées entre 1892 et 2011, c'est-à-dire que certaines sont issues d'études antérieures à l'apparition de la normalisation (Ex : 1972 au Japon, 1982 en France), donc *a priori* entachées d'un niveau d'incertitude supérieur aux autres.

Le seuil de perception olfactive constituant un paramètre fondamental dans la recherche et l'identification des molécules à l'origine de l'odeur d'un mélange gazeux, et compte tenu de toutes les sources potentielles de variabilité qui viennent d'être exposées, la nécessité d'améliorer la fiabilité de ces données apparait donc très clairement. Pour ce faire, une méthodologie de fiabilisation de l'utilisation de ces données est proposée ci-dessous.

2.2 Démarche globale

La méthodologie proposée est un nettoyage qui se décline en cinq étapes :

- Constitution d'une base de données recensant le plus de seuils de perception possible ;
- Élimination des valeurs de seuils de perception issues des articles les moins précis quant à la méthode utilisée pour les mesurer et sur la base de leur densité en nombre de seuils de perception mesurés ;
- Pour chaque molécule, élimination des valeurs aberrantes sur la base de leur loi de distribution et recensement des articles correspondant à ces valeurs aberrantes ;
- Élimination des articles aberrants par rapport au reste du jeu de données en s'appuyant sur le résultat de l'étape 3 ;
- Évaluation de ce nettoyage par comparaison avec une approche expérimentale.

Ces cinq étapes sont décrites dans la suite de ce chapitre. Pour chacune d'entre elles, la méthodologie utilisée est exposée, suivie par l'exposé des résultats de l'application aux données.

Les étapes 2, 3 et 4 conduisent à l'élimination de valeurs de seuils. Pour ces trois étapes, la perte d'information (PI) est présentée sous trois perspectives :

- le nombre d'articles,
- le nombre de seuils de perception total,
- le nombre de molécules renseignées par au moins une valeur de seuil de perception.

L'information la plus importante à surveiller est la dernière citée. En effet, comme cela est expliqué au paragraphe 1.5, la démarche d'identification des molécules responsables de l'odeur dans un mélange rend nécessaire, pour les molécules non renseignées dans la littérature, de leur attribuer une valeur de seuil de perception à partir des molécules de même nature qui sont renseignées. Il est donc important, voire primordial, de préserver un maximum d'informations, c'est-à-dire de conserver un nombre suffisant de molécules renseignées par au moins une valeur de seuil de perception.

Le compromis du nettoyage est donc de conserver une quantité d'informations suffisante tout en éliminant le plus possible d'informations considérées comme non fiables.

La base de données a été constituée sous forme d'un fichier Excel. La méthodologie de fiabilisation de l'utilisation des données a été réalisée à l'aide du logiciel R (version 3.3.1 mise à jour au fur et à mesure du projet) sous la plateforme RStudio.

2.3 Etape 1 : Constitution de la base de données

2.3.1 Définition de la base de données

La base de données constituée regroupe l'intégralité des molécules renseignées par au moins une valeur de seuil de perception dans l'ouvrage de Van Gemert (VAN GEMERT 2011). Seules les molécules pour lesquelles la nature du seuil mesuré (seuil de détection) est clairement précisée dans l'ouvrage ont été prises en compte. Cela correspond à 650 molécules dont les informations sont réparties dans 165 articles. Les molécules constituent les lignes de la base et les articles les colonnes, comme présenté en Figure 14.

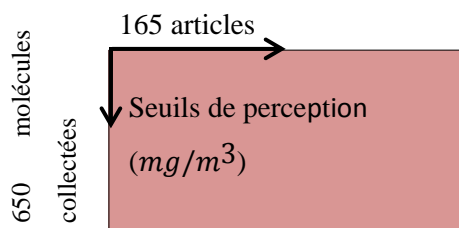


Figure 14 : Format de la base de données des seuils de perception recensés

Cette base de données (BDD1) constitue le support du nettoyage.

Par la suite, BDD1 est convertie en extension « .csv » pour être chargée et analysée à l'aide du logiciel R.

2.3.2 Contenu

La quantité d'information contenue dans BDD1 est présentée sous forme d'une carte de chaleur (« HeatMap ») en Figure 15.

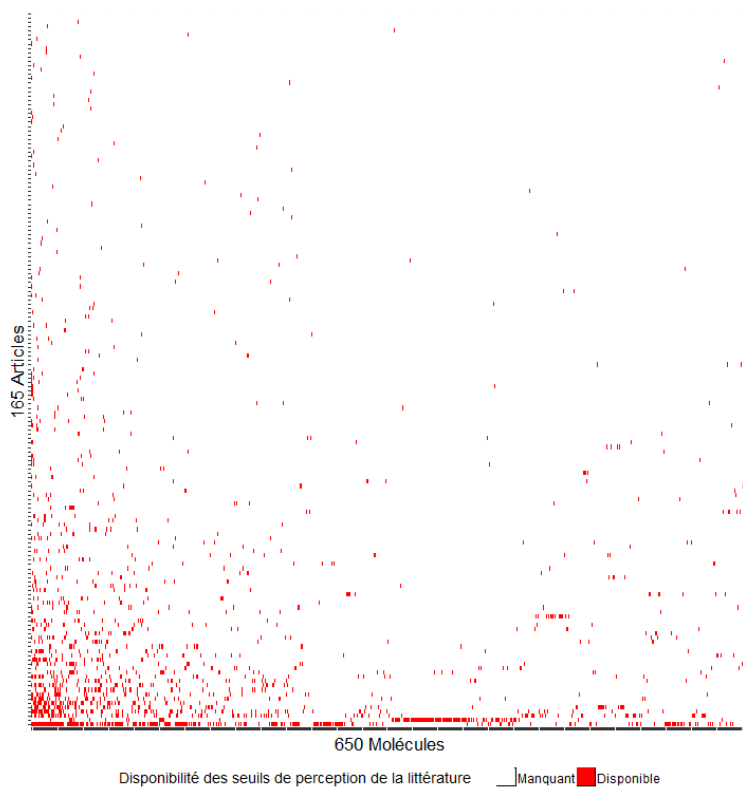


Figure 15 : Quantité d'information disponible dans BDD1 (par ordre décroissant du nombre de valeurs par molécule sur l'axe des abscisses et par ordre décroissant du nombre de valeurs par article sur l'axe des ordonnées)

Sur un plan statistique, cette carte de chaleur illustre le fait que la matrice constituée par les 165 articles et les 650 molécules présente un pourcentage de données manquantes de 98,6%.

La structure des données de la base rend donc impossible l'utilisation de classification statistique et confirme la nécessité de poursuivre avec le nettoyage en plusieurs étapes.

2.4 Étape 2 : Élimination des articles considérés comme les moins fiables

2.4.1 Méthodologie de l'étape 2

Comme présenté en paragraphe 2.2, cette étape consiste à éliminer les articles considérés comme les moins fiables sur la base de leur précision quant à la démarche de mesure utilisée et sur la base de leur densité en nombre de seuils de perception mesurés.

Idéalement, il aurait été pertinent de ne considérer que les articles présentant une évaluation de la répétabilité des mesures effectuées. Le test de Cochran, basé sur l'écart-type de chaque article, aurait alors pu être appliqué [(MILLER 2015) ; (ISO17025 2015)]. Malheureusement, cette information n'est que très rarement disponible pour les articles accessibles (VAN GEMERT 2011) (environ 10% des articles). Ne tenir uniquement compte que des études en faisant mention de cette répétabilité aurait été beaucoup trop restrictif. Par conséquent, le critère de tri qu'il a été choisi d'utiliser est le nombre de mesures de seuils de perception réalisées dans l'étude considérée. Il a en effet été remarqué que plus le nombre de déterminations de seuils de perception dans un article était élevé, plus précise et détaillée était la description de la technique de mesure utilisée, et donc plus fiables pouvaient être considérés les résultats.

Pour limiter la perte d'information globale tout en éliminant les articles considérés comme les moins fiables, un compromis a donc été recherché. Il a consisté à définir le nombre limite de déterminations de seuils de perception dans un article qui pouvait permettre son élimination sans faire perdre plus de 10% de l'information totale. La détermination de cette limite est permise par les résultats du Tableau 4.

Tableau 4 : Distribution des articles en fonction du nombre de valeurs de seuils de perception mesurés

Nombre de seuils de perception par article	Nombre d'articles correspondants	Nombre cumulé d'articles	Nombre de valeurs de seuils de perception	Nombre de valeurs seuils de perception cumulés	Proportion cumulée du total de la base (%)
1	57	57	57	57	4
2	24	81	48	105	7
3	12	93	36	141	9
4	14	107	56	197	13
5	7	114	35	232	15
6	6	120	36	268	18
7	5	125	35	303	20
8	6	131	48	351	23
9	3	134	27	378	25
11	3	137	33	411	27
12	1	138	12	423	28
13	2	140	26	449	30
14	1	141	14	463	31
15	2	143	30	493	33
16	1	144	16	509	34
17	1	145	17	526	35
18	1	146	18	544	36
19	2	148	38	582	39
21	1	149	21	603	40
22	1	150	22	625	42
24	1	151	24	649	43
25	1	152	25	674	45
29	1	153	29	703	47
31	2	155	62	765	51
33	1	156	33	798	53
35	2	158	70	868	58
42	1	159	42	910	61
47	1	160	47	957	64
57	1	161	57	1014	67
77	1	162	77	1091	73
93	1	163	93	1184	79
99	1	164	99	1283	85
220	1	165	220	1503	100
Total	165		1503		

L'examen de la dernière colonne du Tableau 4 montre que pour respecter les 10% de perte maximale d'information totale, nous avons été conduits à choisir d'éliminer les articles contenant strictement moins de quatre valeurs de seuils de perception.

Cette barre de 10% a été choisie pour, comme exposé plus haut, préserver le maximum d'information. Il ne s'agit en effet ici que de la première étape d'élimination de valeurs dans le processus global de nettoyage proposé.

2.4.2 Résultats de l'étape 2

Le détail de la perte d'information sous les trois angles décrits en partie 2.2 est présenté dans le Tableau 5.

Tableau 5 : Bilan de la perte d'information de l'étape 2 du nettoyage

	Étape 2 du nettoyage		
	Avant	Après	Pourcentage de PI
Nombre d'articles	165	72	56%
Nombre de valeurs de seuils de perception	1503	1362	9,4%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	650	629	3,2%

Cette étape conduit à éliminer 56% des articles. Bien que cela représente la moitié des publications, seulement 3% des molécules renseignées par au moins une valeur de seuil de perception a été éliminé. Cette proportion apparaît acceptable. Le manque de fiabilité des sources d'où sont issues ces données justifie leur rejet.

2.5 Étape 3 : Élimination de seuils de perception « aberrants » par molécule

2.5.1 Méthodologie de l'étape 3

2.5.1.1 Définition d'une « valeur de seuil de perception aberrante »

Pour définir ce qu'est une valeur aberrante de seuil de perception, il faut définir la distribution suivie par les valeurs de seuil de perception mesurées pour une même molécule. La taille de l'échantillon recommandée pour cela est de 30 valeurs [(CHAKRAPANI 2011) ; (KAR et RAMALINGAM 2013)]. Malheureusement, aucune des molécules de la BDD1 ne permet de disposer de cette quantité d'information après l'étape 2. Cependant cette taille peut être réduite en fonction du jeu de données (HOGG et TANIS 2005). La fréquence de renseignement des molécules après l'étape 2 est représentée sous forme d'histogramme en Figure 16.

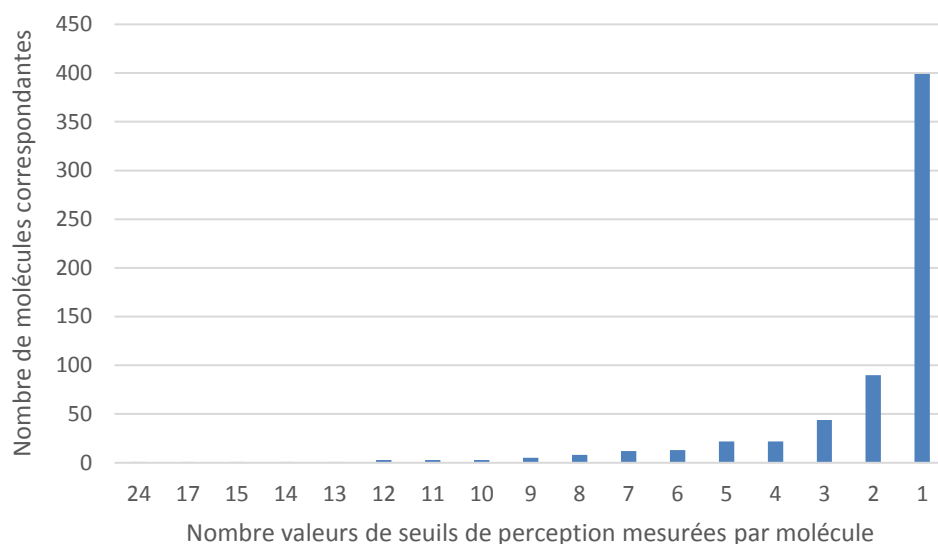


Figure 16 : Fréquence du nombre de valeurs seuils de perception mesurées par molécule

La distribution des valeurs est étudiée pour les 11 molécules les plus renseignées dans la littérature (molécules renseignées de 11 à 24 fois). Dans notre contexte, une molécule analysée dans 11 articles est une molécule dont le comportement a été relativement bien étudié. On peut donc raisonnablement se fier aux résultats qui la concernent. Par exemple, la molécule analysée le plus grand nombre de fois est le n-butanol (butan-1-ol) (24 fois). Ceci s'explique par le fait que c'est la molécule utilisée pour vérifier la normalité olfactive des jurés dans la norme EN 13725 (AFNOR 2003). Cependant, d'un point de vue statistique, ce nombre d'analyses reste faible pour juger de manière fiable de la distribution, comme cela est évoqué plus haut. Il est pourtant nécessaire d'apporter une réponse claire.

À partir d'observations d'experts sur des molécules fréquemment analysées dans la littérature, il est émis l'hypothèse que pour une molécule donnée, les seuils de perception suivent une distribution unimodale et plus précisément normale. La fonction de densité est donc calculée pour les 11 molécules les mieux renseignées. La fonction de densité d'une série de valeurs discrètes est construite sur la base d'une estimation par noyau couramment connue sous le terme de « Kernel Density Estimation » ou « KDE » [(PARZEN 1961) ; (DUONG 2001) ; (VENABLES et RIPLEY 2002)]. Cette fonction est tracée pour ces 11 molécules en Figure 17. La fonction KDE est calculée avec la fonction "geom_density" de R (WICKHAM et al. 2018).

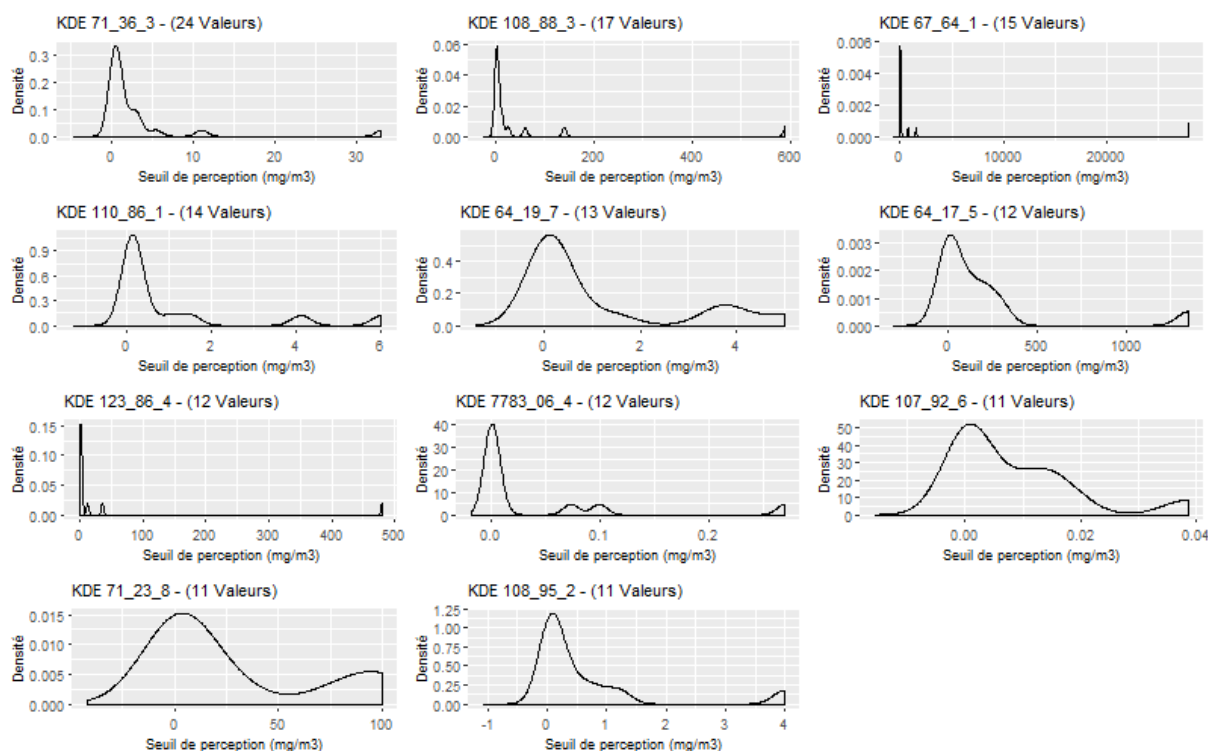


Figure 17 : Distribution des valeurs de seuils de perception des 11 molécules les plus renseignées dans la littérature après l'étape 2

Les molécules correspondant aux numéros CAS sont présentées en Tableau 6.

Tableau 6 : Molécules renseignées par un minimum de 10 valeurs de seuil de perception

N° CAS	Molécules	Famille chimique
71-36-3	butan-1-ol	Alcools
108-88-3	toluene	Aromatiques
67-64-1	propan-2-one	Cétones
110-86-1	pyridine	Azotés
64-19-7	acetic acid	Acides
64-17-5	ethanol	Alcools
123-86-4	butyl acetate	Esters
7783-06-4	sulfane	Soufrés
107-92-6	butanoic acid	Acides
71-23-8	propan-1-ol	Alcools
108-95-2	phenol	Aromatiques

La Figure 17 permet de constater que les valeurs de seuils de perception peuvent être approchées par une densité unimodale, présentant ainsi un mode significatif. Néanmoins, en raison de certaines valeurs marginales, elles ne semblent pas toujours suivre une distribution normale. Numériquement, il existe plusieurs tests pour tester la normalité [(MORICE 1972) ; (SAPORTA 2011)]. Cependant, ces tests tendent fréquemment à rejeter l'hypothèse de la loi normale dans le cas d'échantillons de taille importante ($\gg 30$) et à admettre la normalité dans le cas d'échantillons de petite taille (< 30) (MORICE 1972). Parmi ceux-ci, le test de Shapiro Wilk est l'un des plus utilisés en raison de sa puissance qui

néanmoins est moindre dans le cas d'échantillon de faible taille (<30) (RAZALI et WAH 2011). C'est pourquoi nous avons préféré utiliser comme indicateur le coefficient de corrélation du diagramme Quantile-Quantile (Q-Q) de chacune des 11 molécules. Ce diagramme permet d'observer graphiquement l'ajustement d'une distribution observée à une distribution théorique (GOLDFARB et PARDOUX 2005). Ici, la distribution théorique testée est « normale » et le diagramme Quantile-Quantile est tracé avec la fonction "qqnorm" de R. De par la définition du diagramme Quantile-Quantile, plus le coefficient de corrélation de la courbe obtenue est grand, plus cela signifie que les valeurs suivent une distribution normale. Prenons l'exemple du diagramme Quantile-Quantile du n-butanol (butan-1-ol) (Figure 18).

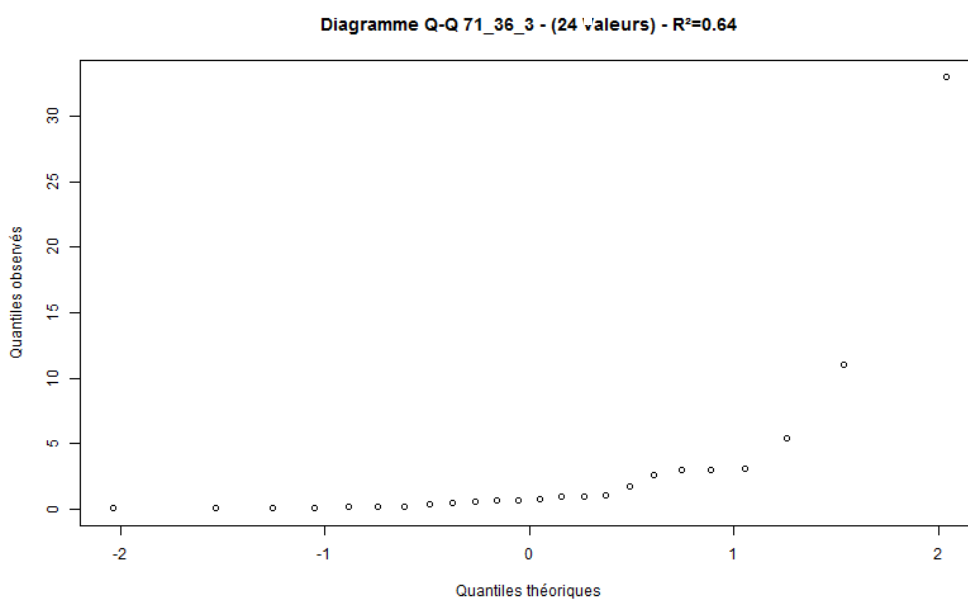


Figure 18 : Diagramme Q-Q du n-butanol (CAS=71-36-3)

Le nuage de points semble davantage signifier une asymétrie à droite qu'une loi normale. Cependant, au regard de la représentation de la fonction KDE du n-butanol en Figure 17, cette asymétrie s'expliquerait par des valeurs ponctuelles ayant un ordre de grandeur incohérent avec le reste des données qui interfèreraient dans le résultat. Le coefficient de corrélation (CC) entre l'axe des abscisses et l'axe des ordonnées est de 0,64. Une valeur plus élevée aurait été préférable pour valider directement le caractère normal de la distribution, mais cette valeur suffit pour considérer la distribution comme normale au vu de ce qui a été dit précédemment. Le coefficient de corrélation du diagramme Quantile-Quantile des 11 molécules est calculé et représenté en Figure 19.

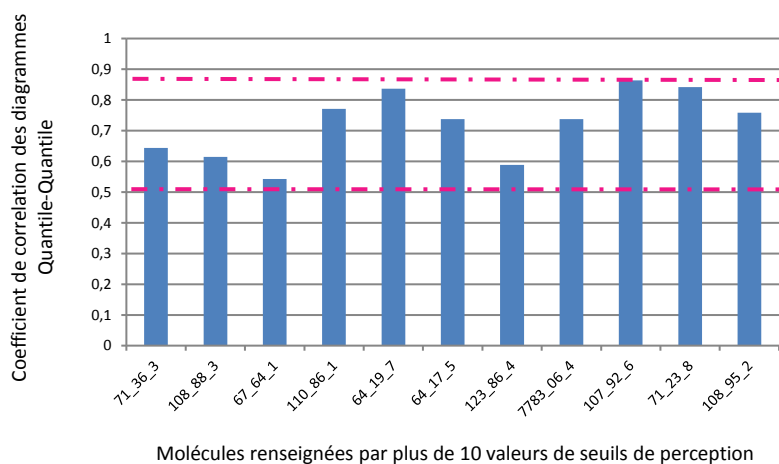


Figure 19 : Coefficient de corrélation (CC) des diagrammes Quantile-Quantile (Q-Q) des molécules dont le seuil de perception est renseigné par strictement plus de 10 articles après l'étape 2 du nettoyage

Les coefficients de corrélation se situent entre 0,5 et 0,9. De la même façon que pour l'étude du n-butanol, si l'on considère les valeurs ponctuelles ayant un ordre de grandeur incohérent avec le reste des données, alors cela encourage à valider l'hypothèse que les seuils de perception d'une même molécule suivent une distribution normale.

Dans l'étape qui suit, les valeurs aberrantes sont identifiées à l'aide d'un test valide sous la condition d'une distribution normale des valeurs.

2.5.1.2 Élimination des valeurs de seuils de perception aberrantes selon les critères d'une distribution normale

Pour identifier les valeurs de seuils de perception aberrantes pour chaque molécule, le test de reproductibilité inter-laboratoire de Grubbs [(MILLER 2015) ; (ISO17025 2015)] a été choisi. Le niveau de confiance sélectionné est de 95%. Pour identifier les valeurs aberrantes, ce test est préféré à l'utilisation de la boîte à moustache car cette dernière est avant tout un outil de visualisation graphique de la dispersion des données. De plus, le test de Grubbs peut être appliqué à partir d'une série de données comportant trois valeurs (ASTM International 1994) tandis que la boîte à moustache n'est pas pertinente dans un tel cas. Même si les deux tests non paramétriques de Grubbs et de Dixon sont adaptés aux cas où l'on dispose d'un nombre peu élevé de valeurs (PILLET, PILLET, et ALBEROLA 2018), le test de Grubbs est préféré au test de Dixon qui lui apparaît plus pertinent dans un cas d'analyse intra-laboratoire. Étant donné que le test de Grubbs n'est applicable qu'à partir d'une série de données de trois valeurs (ASTM International 1994), seules les seuils de perception des molécules dont celui-ci a été mesuré au moins trois fois peuvent être comparés au reste de la base de données à cette étape. Ces seuils sont nommés « seuils de perception testés » dans la suite du document. De ce fait, à ce stade, les valeurs de

seuils de perception des molécules renseignées strictement moins de deux fois ne sont pas comparées au reste de la base.

Un exemple de cette situation est présenté dans le Tableau 7.

Tableau 7 : Exemple 1 : molécules renseignées une seule fois (valeurs de seuils de perception en mg/m^3)

Famille	No CAS	Molécules	(LINDELL 1991)	(VON RANSON et BELITZ 1992)	(NAGATA 2003)	Nombre de mesures par molécule
Aldéhydes	112-31-2	decanal	0,051	0,063	0,026	3
Aldéhydes	112-44-7	undecanal	–	0,14	–	1
Aldéhydes	112-54-9	dodecanal	–	0,033	–	1

Dans ce tableau, les valeurs de seuils de perception mesurées par Von Ranson et Belitz peuvent être comparées à d'autres valeurs de la base pour la molécule de décanal. Cependant, pour l'undécanal et le dodécanal cette comparaison n'est pas possible. Ainsi, à cette étape 3, la valeur est conservée sans justification. Un autre exemple est présenté dans le Tableau 8.

Tableau 8 : Exemple 2 : molécules renseignées une seule fois (valeurs de seuils de perception en mg/m^3)

Famille	No CAS	Molécules	(MAY 1966)	(HELLMAN et SMALL 1974)	(DRAVNIKS 1974)	(DON 1986)	(NAGY 1991)	(NAGATA 2003)	Nombre de mesures par molécule
Alcools	109-86-4	2-methoxyethanol	190	0,3	–	–	–	–	2
Cétones	96-22-0	pentan-3-one	33	–	3	–	–	–	2
Chlorés	107-06-2	1,2-dichloroethane	450	25	–	–	–	–	2
Chlorés	71-55-6	1, 1,1-trichloroethane	2100	–	–	5,3	–	–	2
Esters	107-31-3	Methyl formate	5000	–	–	–	–	325	2
Esters	592-84-7	butyl formate	70	–	–	–	–	0,37	2
Ethers	109-99-9	oxolane	90	–	–	–	18	–	2
Alcanes	111-65-9	octane	710	–	–	–	61,8	8	3
Alcanes	142-82-5	heptane	930	–	–	–	110	2,7	3
Esters	110-19-0	2-methylpropyl acetate	17	1,7	–	–	–	0,038	3

Dans ce cas, les mesures de seuils de perception de May sont comparées à d'autres valeurs de la base pour l'octane, l'heptane et l'acétate d'isobutyle (2-methylpropyl acetate), mais pas pour les autres molécules pour lesquelles on ne dispose que de deux mesures de seuils de perception. Or, pour ces molécules, il existe une hétérogénéité importante entre les valeurs de l'article de May et celles des autres articles. Il est donc nécessaire de disposer d'une information sur la fiabilité de ces valeurs malgré leur faible nombre.

C'est dans cet objectif que la présente étape de nettoyage est complétée par une quatrième étape portant sur une détection des articles jugés « aberrants ». Pour cela, on recensera le nombre de fois où chaque article est éliminé à l'étape 3 pour réaliser l'étape 4.

2.5.2 Résultats de l'étape 3

Les résultats de la perte d'information à l'étape 3 sont présentés dans le Tableau 9.

Tableau 9 : Bilan de la perte d'information à l'étape 3 du nettoyage

	Etape 3 du nettoyage		
	Avant	Après	Pourcentage de PI
Nombre d'articles	72	72	0%
Nombre de seuils de perception	1362	1233	9,5%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	629	629	0%

Cette étape aboutit à une proportion d'élimination de valeurs de seuils de perception relativement faible (<10%) et à une absence de perte d'information pour les deux autres critères. Le détail de cette perte d'information est présenté dans le Tableau 10.

Tableau 10 : Élimination de valeurs de seuils de perception : détail de la perte d'information

Nombre de valeurs éliminées par molécule	Nombre de molécules concernées	Nombre total des valeurs éliminées
1	46	46
2	21	42
3	6	18
4	4	16
7	1	7
TOTAL	78	129

Pour chacune des 78 molécules, au moins une valeur de seuil de perception est éliminée. Parmi ces molécules, 11 ont subi une élimination supérieure à 2 valeurs. Celle ayant le plus de valeurs éliminées (7 valeurs) est le n-butanol (butan-1-ol), qui, comme expliqué plus haut, est la plus renseignée dans la littérature (38 valeurs de seuil de perception initialement, 24 après l'étape 2).

En moyenne 1,65 valeurs ont été éliminées sur ces 78 molécules. Ce faible taux d'élimination tend à nouveau à confirmer la distribution normale des valeurs.

Les seuils de perception éliminés lors de cette étape appartiennent souvent aux mêmes articles. Cette observation a permis de définir l'approche mise en œuvre dans l'étape 4.

2.6 Étape 4 : Élimination des articles « aberrants »

2.6.1 Méthodologie

La démarche utilisée consiste à comptabiliser le nombre de fois où des valeurs de seuils de perception mentionnées dans un article ont été éliminées lors de l'étape 3. Pour chaque article « j », une fréquence d'exclusion (FE) est calculée (Équation 4). Cette fréquence traduit la pertinence de l'article.

$$FE_j = \frac{N_{SP\ exclus}^j}{N_{SP\ testés}^j} \times 100$$

FE_j : Fréquence d'exclusion de l'article j
 $N_{SP\ exclus}^j$: Nombre de seuils de perception de l'article j exclus par le test de Grubbs à l'étape 3
 $N_{SP\ testés}^j$: Nombre de seuils de perception de l'article j testés

Équation 4 : Définition de la fréquence d'exclusion (FE) d'un article « j »

En parallèle, il faut prendre en compte le fait que les articles n'ont pas été comparés le même nombre de fois. En effet, les valeurs de certains articles ont été comparées en forte proportion alors que d'autres n'ont jamais été comparées.

Plus les valeurs de seuils de perception d'un article ont fait l'objet de comparaisons, plus la fréquence d'exclusion de cet article est robuste. Un taux de comparaison (TC) est donc calculé pour chaque article « j » (Équation 5).

$$TC_j = \frac{N_{SP\ testés}^j}{N_{SP\ mesurés}^j} \times 100$$

TC_j : Taux de comparaison de l'article j
 $N_{SP\ testés}^j$: Nombre de seuils de perception de l'article j testés à l'étape 3
 $N_{SP\ mesurés}^j$: Nombre de seuils de perception mesurés dans l'article j

Équation 5 : Définition du taux de comparaison (TC) d'un article « j »

Il a été choisi de conserver les articles dont strictement plus de 2/5^{ème} des valeurs de seuils de perception avaient été comparées à d'autres valeurs de seuils de la base de données (TC>40%). Ce choix est motivé par la volonté de préférer conserver de l'information robuste et de qualité plutôt qu'une grande quantité d'information sujette à caution. En effet, un taux de comparaison strictement supérieur à 40% évite, du point de vue de l'expert, la présence de valeurs incohérentes dans la base de données.

Les derniers articles satisfaisant à notre critère sont ensuite classés. La classification est appliquée sur la base des fréquences d'exclusion (FE) de ces articles et correspond à une Classification Ascendante Hiérarchique (CAH) basée sur l'indice de distance de Ward (WARD 1963). Cette classification permet de séparer statistiquement les articles en fonction, ici, des valeurs de fréquences d'exclusion. La fonction R utilisée est « hclust » du package « stats » avec comme paramètre « method » l'indice « ward.D2 » qui correspond à l'indice décrit par Ward (WARD 1963). Le dendrogramme et le graphique de l'évolution du coefficient de corrélation semi-partiel (SPRSQ ou « *Semi-partial R-squared* ») sont utilisés pour déterminer le nombre de groupes [(LEBARBIER et MARY-HUARD 2008) ; (TUFFERY 2008) ; (TUFFERY 2011)]. En effet, ces deux graphiques apportent de l'information sur l'inertie des classes dont l'observation est le critère de sélection d'une classification pertinente. Le premier permet d'observer le saut d'inertie le plus important (LARMARANGE 2018). Le second mesure la perte d'inertie interclasse et sert à repérer la chute la plus importante du coefficient de corrélation semi-partiel (SPRSQ). À la fin, le groupe contenant la fréquence d'exclusion minimale est retenu.

2.6.2 Résultats de l'étape 4

La première partie de la démarche est de ne conserver que les articles dont le taux de comparaison est strictement supérieur à 40% (Annexe 1). Le nombre d'articles éliminés est de 15 (Tableau 11).

Tableau 11 : Articles éliminés (TC ≤ 40%)

Articles	Nombre de seuils de perception aberrants selon le test de Grubbs	Nombre de seuils de perception testés par le test de Grubbs	Total des seuils de perception mesurés dans l'article	FE (%)	TC (%)
(EYKMAN 1927)	0	0	4	0	0
(FORS et OLOFSSON 1985)	0	0	13	0	0
(BERGER et al. 1985)	0	0	11	0	0
(MAARSE, NIJSSEN, et ENGELINO 1988)	0	0	4	0	0
(LASKA 2010)	0	0	6	0	0
(SAVENHED, BOREN, et GRIMVALL 1985)	0	1	8	0	12
(WUNSCHÉ et al. 1995)	0	1	7	0	14
(HUIJER 1917)	0	1	5	0	20
(KRAFT et POPAJ 2004)	0	1	5	0	20
(PATTE et PUNTER 1979)	0	2	8	0	25
(BAYDAR, PETRZILKA, et SCHOTT 1993)	0	1	4	0	25
(STUIVER 1958)	0	10	25	0	40
(HELLMAN et SMALL 1974)	1	36	93	3	39
(VON RANSON et BELITZ 1992)	3	13	99	23	13
(MOLHAVE et al. 2000)	1	2	5	50	40

Le Tableau 11 permet d'observer qu'en l'absence de tri selon le taux de comparaison (TC), de nombreux articles dont la fréquence d'exclusion (FE) est nulle aurait été conservés sans justification de fiabilité, comme par exemple dans le cas de « Berger et al. » (BERGER et al. 1985).

La perte d'information est présentée dans le Tableau 12.

Tableau 12 : Bilan de la perte d'information après élimination des articles dont le TC ≤ 40%

	Étape 4 du nettoyage après sélection des articles selon le TC		
	Avant	Après	Pourcentage de PI
Nombre d'articles	72	57	20,8%
Nombre de seuils de perception	1233	941	23,7%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	629	435	30,8%

Les fréquences d'exclusion (FE) des articles restants sont utilisées pour les classer conformément à ce qui a été expliqué au paragraphe 2.6.1. Le dendrogramme de la classification ascendante hiérarchique (CAH) appliquée est présenté en Figure 20.

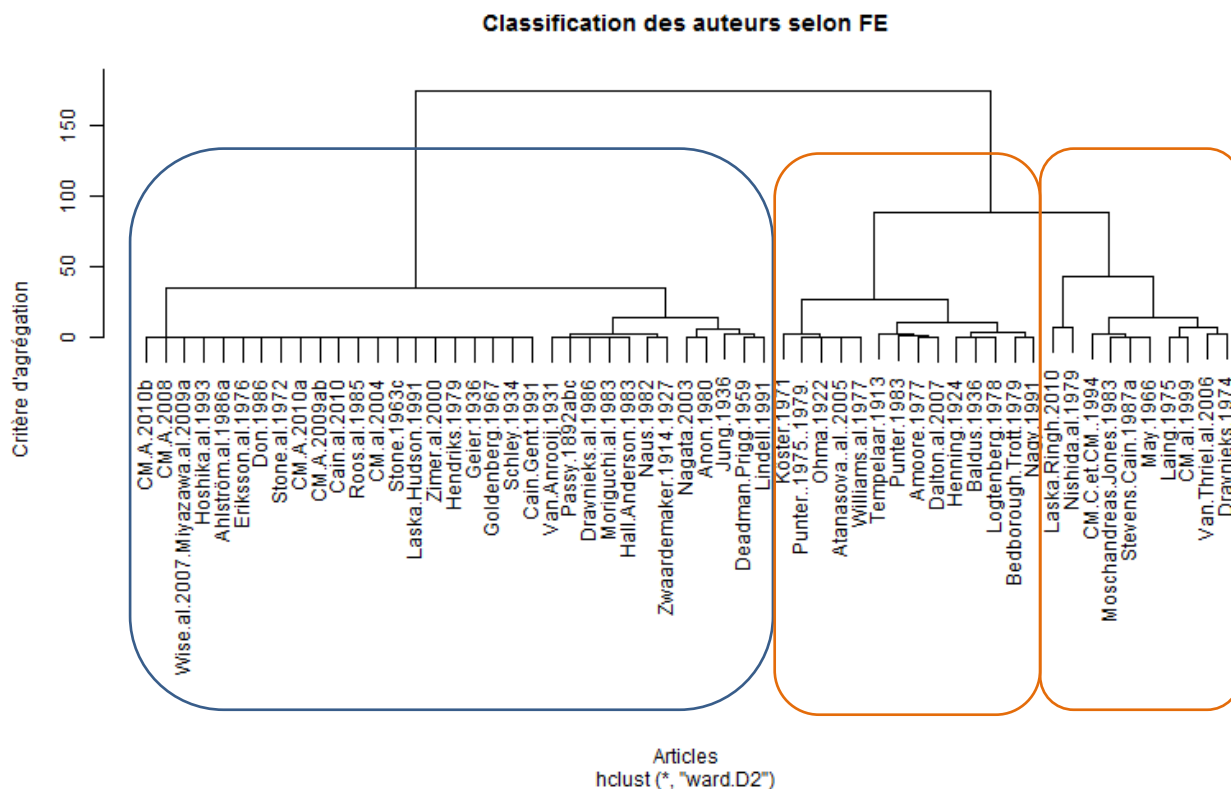


Figure 20 : Dendrogramme de la CAH appliquée sur les FE des articles ayant un TC strictement supérieur à 40%

La Figure 20 met en relief une forte chute du critère d'agrégation de 180 à 50. Cela permet de distinguer un premier groupe (encadré en bleu). L'organisation des données au sein de ce groupe est ensuite stable. La deuxième branche divise ensuite le reste des données (encadrés en orange) avec une chute du critère d'agrégation certes notable mais moins importante que précédemment. Une division plus poussée sur cette deuxième branche génèrerait une division du premier groupe, ce qui ne serait plus optimal. Ce dendrogramme conduit à distinguer trois classes.

Cette figure est complétée par le graphique de perte d'inertie (Figure 21).

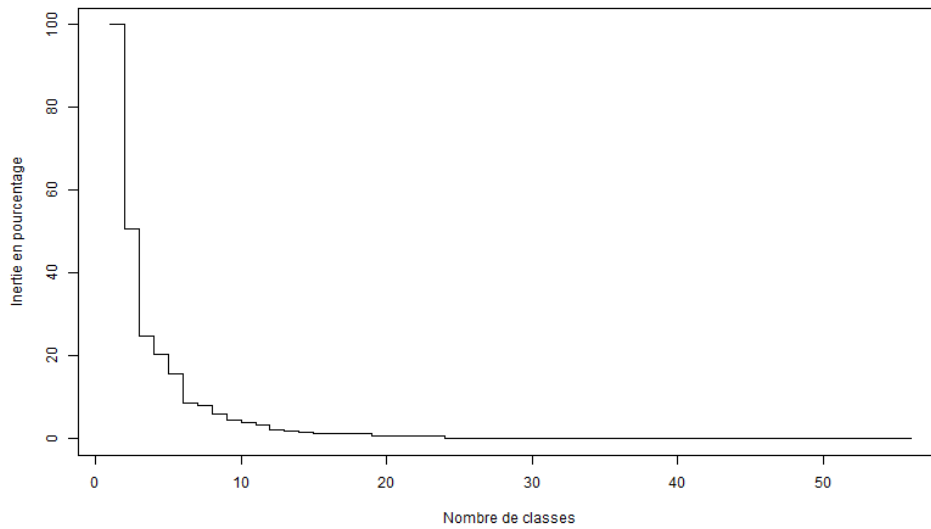


Figure 21 : Évolution de la perte d'inertie (en pourcentage) en fonction du nombre de classes

La Figure 21 montre que l'inertie se stabilise et est régulière à partir du deuxième saut. Cela confirme que le nombre de classes approprié est de trois.

Un troisième graphe d'aide à la décision du nombre de composantes est le graphique de l'évolution du coefficient de corrélation semi-partiel (SPRSQ). Il est présenté en Figure 22.

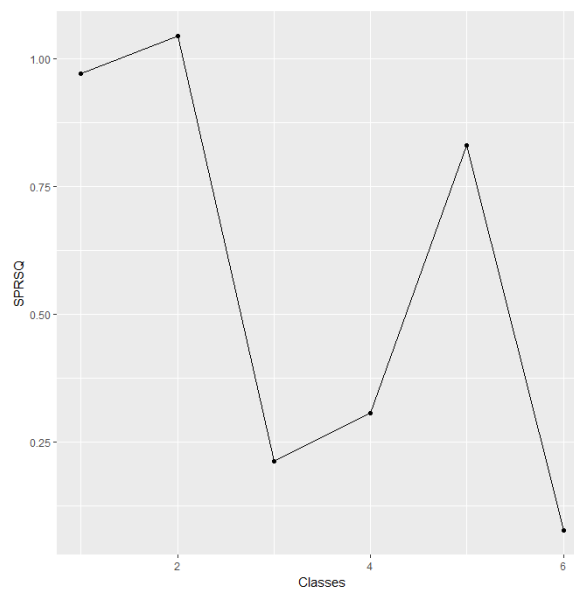


Figure 22 : Évolution du SPRSQ en fonction du nombre de classes

Cette dernière figure confirme le choix d'appliquer trois classes puisque la première chute importante du coefficient de corrélation semi-partiel (SPRSQ) correspond au passage de 2 à 3 classes. Le résultat de cette classification est présenté en Annexe 2.

Le groupe contenant la fréquence d'exclusion (FE) minimale est le groupe 1 ($FE \leq 12$). Il contient 33 articles. C'est ce groupe d'articles qui est conservé.

Le résultat de perte d'information suite à cette étape est présenté dans le Tableau 13.

Tableau 13 : Bilan de la perte d'information après élimination des articles (CAH sur les FE)

	Étape 4 du nettoyage après classification des articles		
	Avant	Après	Pourcentage de PI
Nombre d'articles	57	33	42,1%
Nombre de seuils de perception	941	614	34,7%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	435	338	22,3%

Le bilan de la perte d'information global de la quatrième étape est présenté en Tableau 14.

Tableau 14 : Bilan de la perte d'information de l'étape 4 du nettoyage

	Étape 4 du nettoyage - Bilan		
	Avant	Après	Pourcentage de PI
Nombre d'articles	72	33	54%
Nombre de seuils de perception	1233	614	50%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	629	338	46%

La perte d'information lors de cette étape apparait comme relativement drastique par rapport à celles constatées lors des précédentes étapes. Cette étape aboutit en effet à la perte de globalement la moitié du jeu de données. Il reste toutefois un nombre très raisonnable de molécules renseignées par au moins une valeur de seuil de perception. Cette information est répartie sur 33 articles.

Une fois de plus, l'acceptation d'un taux de perte d'information aussi élevé est justifiée par la volonté de privilégier la fiabilité des données plutôt que la quantité.

2.7 Étape 5 : Évaluation de la qualité du nettoyage à partir d'un référentiel expérimental de seuils de perception

2.7.1 Méthodologie de l'étape 5

2.7.1.1 Définition du référentiel

Le processus de nettoyage dont nous venons d'exposer les différentes étapes résulte d'une succession de choix et de traitements statistiques à partir de données de seuils de perception disponibles dans la littérature. Un autre moyen d'obtenir ces données est la mesure, mais c'est un moyen coûteux et

chronophage. Ces mesures permettent néanmoins d'obtenir une information maîtrisée quant à l'incertitude et donc de disposer d'un référentiel expérimental de comparaison. L'objectif visé ici est de comparer le résultat du nettoyage statistique à des mesures afin de pouvoir statuer sur l'efficacité du nettoyage statistique sur l'ensemble de la base.

Le seuil de perception de 45 molécules a donc été mesuré par Olentica (Annexe 3). Les résultats consensus sont présentés dans le Tableau 15.

Tableau 15 : Mesures des seuils de perception pour 45 molécules du référentiel

N° CAS	Molécules	Mode des valeurs	Ecart-type	Nombre de mesures
108-88-3	toluene	1,22E+00	1,48E-01	12
71-36-3	butan-1-ol	5,21E-02	6,74E-02	11
123-86-4	butyl acetate	5,13E-02	2,01E-02	9
120-92-3	cyclopentanone	9,11E-01	2,43E-01	9
470-82-6	2,2,4-trimethyl-3-oxabicyclo[2.2.2]octane	1,44E-03	5,64E-03	9
66-25-1	hexanal	3,25E-03	2,59E-03	4
623-42-7	méthyl butanoate	3,32E-02	8,13E-03	7
75-08-1	éthanethiol	3,24E-05	9,64E-06	4
5989-27-5	(4R)-1-methyl-4-prop-1-en-2-ylcyclohexene (D-Limonène)	8,29E-02	1,90E-02	8
121-44-8	N,N-diethylethanamine	4,01E-02	1,76E-01	12
431-03-8	butane-2,3-dione	1,17E-04	2,56E-05	6
104-76-7	2-ethylhexan-1-ol	3,81E-01	9,14E-02	4
140-11-4	benzyl acetate	1,73E-01	8,20E-02	5
80-56-8	4,6,6-trimethylbicyclo[3.1.1]hept-3-ene	1,92E-02	8,61E-04	3
123-05-7	2-ethylhexanal	7,73E-02	8,90E-02	7
99-85-4	1-methyl-4-propan-2-ylcyclohexa-1,4-diene	4,69E-01	2,87E-02	3
112-31-2	decanal	2,81E-03	4,09E-03	8
106-21-8	3,7-dimethyloctan-1-ol	4,59E-03	4,41E-03	7
591-78-6	hexan-2-one	6,03E-01	5,42E-01	6
589-38-8	hexan-3-one	1,09E-01	3,55E-02	6
60-12-8	2-phenylethanol	1,53E-02	4,49E-03	5
110-54-3	hexane	1,24E+01	1,12E+01	6
592-41-6	hex-1-ene	5,22E-04	2,17E-03	10
1634-04-4	2-methoxy-2-methylpropane (MTBE)	2,50E-02	1,10E-02	9
64-17-5	ethanol	9,36E-02	1,22E-01	14
98-86-2	1-phenylethanone	1,30E-02	3,77E-03	10
105-54-4	ethyl butanoate	5,78E-04	8,48E-05	8
624-24-8	methyl pentanoate	1,01E-02	1,76E-03	8
590-86-3	3-methylbutanal	2,20E-04	2,07E-04	6
75-18-3	Methylsulfanylmethane (DMS)	4,00E-03	1,51E-03	7
67-64-1	propan-2-one	7,75E+00	6,68E+00	12
78-93-3	butan-2-one	2,54E+00	4,93E-01	6
141-78-6	ethyl acetate	2,91E+00	9,40E-01	6
111-66-0	oct-1-ene	7,18E-03	2,08E-03	6
109-99-9	Oxolane (THF)	2,26E+00	5,68E-01	6
60-29-7	ethoxyethane	1,00E-01	3,21E-02	6
96-17-3	2-methylbutanal	1,68E-03	1,52E-03	7
592-76-7	hept-1-ene	1,43E-01	3,82E-02	6
105-37-3	ethyl propanoate	2,19E-02	9,09E-03	7
108-10-1	4-methylpentan-2-one	7,21E-01	9,84E-02	6
100-42-5	styrene	5,92E-02	1,42E-02	6
110-83-8	cyclohexene	2,35E-01	4,52E-02	6
95-47-6	1,2-xylene	1,49E+00	2,33E-01	6
110-43-0	heptan-2-one	7,81E-02	2,15E-02	6
108-83-8	2,6-dimethylheptan-4-one	4,01E-01	1,21E-01	6

Pour chacune de ces molécules le seuil de perception a été mesuré selon la norme EN 13725 entre 3 et 14 fois. Le protocole de génération de l'échantillon gazeux est présenté en Annexe 4.

Le choix de ces 45 molécules a été réalisé pour assurer la représentativité non seulement des différentes familles chimiques susceptibles d'être rencontrées, mais aussi des différents ordres de grandeur des seuils de perception. Parmi ces 45 molécules, 41 sont renseignées dans la littérature au moins une fois. La raison pour laquelle le seuil de perception de molécules non renseignées dans la littérature a été mesuré est expliquée au chapitre 4. Ces mesures constituent notre référentiel expérimental. La première chose à réaliser est une analyse de ce référentiel.

2.7.1.2 Analyse du référentiel

Le but de cette étape est d'évaluer la qualité des mesures de seuils de perception constituant le référentiel ainsi que sa variabilité globale.

L'ensemble des valeurs de seuils de perception d'une même molécule du référentiel est nommé « variable ». Le Tableau 15 permet de constater que ces variables ont des ordres de grandeur et des écart-types différents.

De ce fait, afin de pouvoir analyser ce référentiel, une étape de standardisation par molécule est appliquée. Les paramètres choisis pour cette standardisation sont la moyenne et l'écart-type de chaque variable. Cela est justifié par le caractère unimodal et relativement symétrique de la distribution des variables. Le cas échéant, il est possible d'appliquer d'autres paramètres tels que la médiane à la place de la moyenne et l'intervalle interquartile divisé par deux à la place de l'écart-type (GRASLAND 2000). L'utilisation de cette standardisation permet de comparer les molécules sur une même échelle et d'estimer la variation globale. En effet, la standardisation fait disparaître l'unité de mesure de la variable d'origine. Cette dernière est alors mesurée en écart-type et exprime l'écart d'un élément de la distribution à la moyenne. Chacune des variables standardisées est comprise dans un intervalle [-3 ; +3] étant donnée la normalité des variables (Minitab 18 2017). Il est alors possible de repérer les données aberrantes, généralement situées en dehors de l'intervalle [-2 ; +2] (FRANCOIS et GRASLAND 2000). Pour identifier ces valeurs statistiquement aberrantes, un test adapté aux mesures intra-laboratoires et à une distribution normale des variables a été choisi (EDUSCOL 2016) : Il s'agit du test de Dixon [(DEAN et DIXON 1951) ; (CHAPUT et al. 2013)]. Celui-ci est applicable car, comme présenté au paragraphe 2.5.1, les seuils de perception d'une même molécule tendent à suivre une distribution normale. Le choix du risque est fixé à 5%. Un bilan du pourcentage de valeurs éliminées permet de juger du niveau de variabilité du référentiel et de sa fiabilité.

En parallèle, pour évaluer la variabilité du référentiel, la répétabilité des mesures non standardisées est mesurée pour chaque molécule (INRS 2000). Un intervalle de confiance est donc obtenu pour chacune des molécules. Cependant, les ordres de grandeurs sont tellement différents qu'un intervalle de confiance moyen ne peut pas être calculé directement.

Pour cette raison, il a été choisi de travailler sur le jeu de données standardisé par la variance et de mesurer la dispersion de chacune des variables standardisées par un écart absolu moyen (EAM). En effet, la répétabilité étant basée sur l'écart-type des données, il n'y aurait pas d'intérêt à la calculer sur des données standardisées puisqu'elle ne dépendrait que du nombre de valeurs. Cet écart absolu moyen correspond à la moyenne de la valeur absolue des écarts à la moyenne de chacune des molécules [(FRANCOIS et GRASLAND 2000) ; (POLISANO 2018)]. Cet indicateur a été préféré à l'intervalle interquartile ou à l'étendue de chaque série divisée par 2 car le premier nécessitait un nombre important de valeurs pour obtenir un résultat représentatif et le deuxième ne donnait que deux informations sur les séries de données (le minimum et le maximum) et n'était pas représentatif de la série de données (FRANCOIS et GRASLAND 2000).

In fine, la moyenne et l'écart-type des écarts absolus moyens individuels sont calculés pour obtenir un écart à la moyenne moyen du référentiel et sa distribution.

2.7.1.3 Évaluation du nettoyage par molécule

Afin d'analyser plus précisément l'effet du nettoyage, une analyse est réalisée par molécule. Pour cette analyse, aucune standardisation n'est utilisée. Pour chaque molécule, la valeur représentative utilisée pour la suite du projet est le mode¹ des fonctions de densité des séries de valeurs. C'est pourquoi le mode des différents types de données est comparé par molécule. Six valeurs sont définies (Tableau 16).

Tableau 16 : Définition des valeurs calculées par molécule pour comparaison des jeux de données

X_1 :	mode des valeurs de SP brutes de la littérature
X_2 :	mode des valeurs de SP de la littérature restantes après l'application de la méthodologie de nettoyage
X_3 :	mode des valeurs de SP expérimentales
$Z_1 =$	$ X_1 - X_3 $
$Z_2 =$	$ X_2 - X_3 $
$Re =$	Répétabilité de la mesure de X_3

Pour chaque molécule du référentiel, la valeur est considérée comme correcte si Z_1 et Z_2 sont inférieures à la répétabilité (Re). La répétabilité reflète la variabilité de la mesure du seuil de perception d'une molécule. Une valeur élevée peut traduire un comportement atypique de la molécule rendant difficile et donc variable la mesure de son seuil de perception. Par exemple, le caractère anesthésiant d'une molécule pour l'appareil olfactif, comme c'est le cas pour l'eucalyptol.

Le nettoyage appliqué sera néanmoins considéré comme efficace si au moins l'une des deux conditions suivantes est respectée :

- $Z_2 < Z_1$

¹ Valeur d'un caractère quantitatif discret ayant la fréquence la plus élevée (LAROUSSE 2018)

- $Z_2 < Re$

Le nombre de cas où le nettoyage est efficace est ensuite comptabilisé.

2.7.2 Résultats de l'étape 5

2.7.2.1 Résultats de l'analyse du référentiel

Comme cela est présenté au paragraphe 2.7.1, les données du référentiel sont standardisées. Une HeatMap de ces valeurs est représentée (Figure 23).

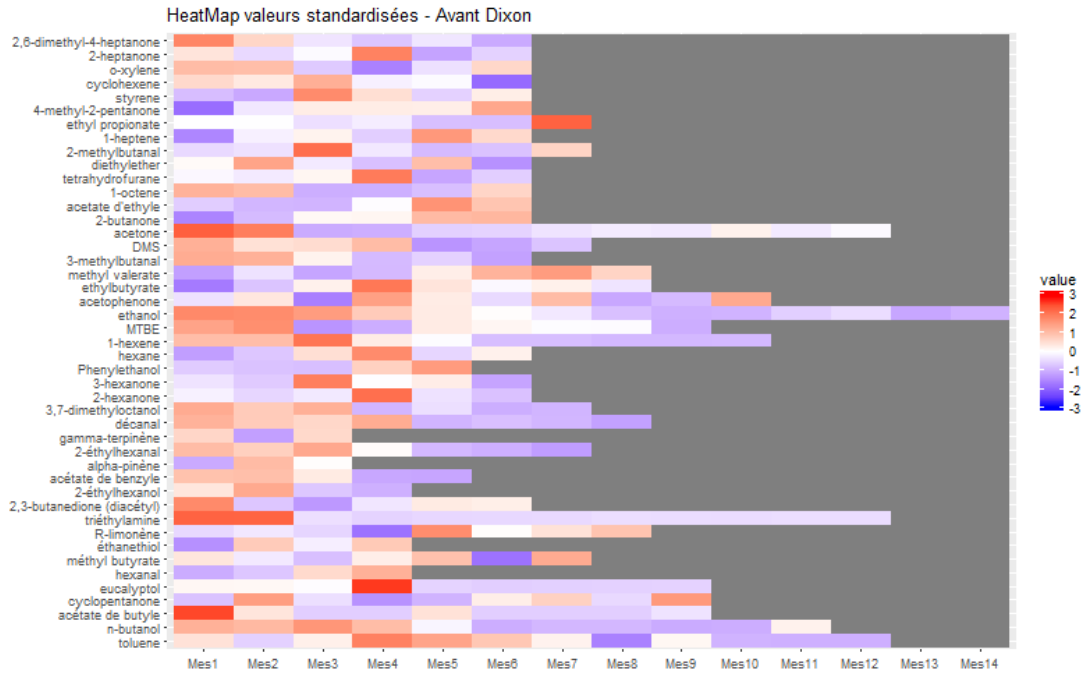


Figure 23 : HeatMap des valeurs de seuils de perception du référentiel standardisées par molécule

Graphiquement, il est possible de détecter les valeurs aberrantes situées dans des zones « rouge vif » ou « bleu vif ». Cela correspondant à des zones extrêmes. Statistiquement, le test de Dixon est appliqué sur les données standardisées. La HeatMap des valeurs restantes est représentée en Figure 24.

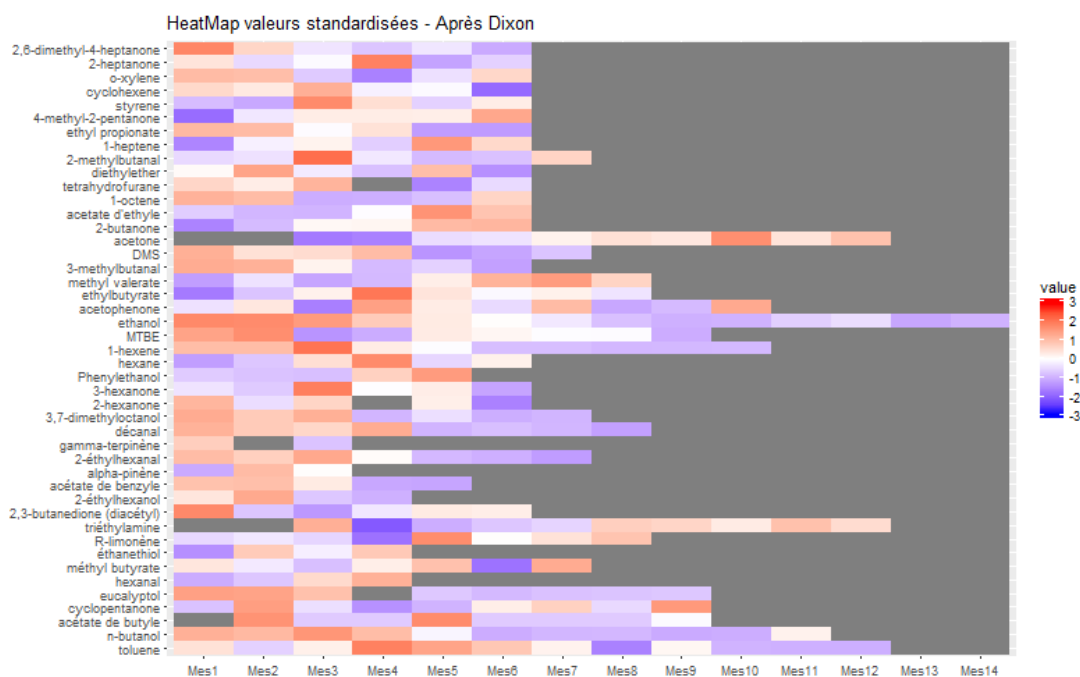


Figure 24 : HeatMap des valeurs de seuils de perception du référentiel standardisées par molécules après application du test de Dixon

Les valeurs des zones extrêmes ont été éliminées par le test de Dixon. À présent, les données sont comprises dans l'intervalle $[-2,22 ; 1,82]$. Le test de Dixon appliqué aux données non standardisées mène à la sélection des mêmes valeurs aberrantes à l'exception d'une valeur de l'éthyl butyrate (ethyl butanoate) (Annexe 5).

Le détail des valeurs éliminées est présenté dans le Tableau 17.

Tableau 17 : Valeurs de seuils de perception (SP) éliminées par le test de Dixon

N° CAS	Molécules	SP exclus par Dixon	SP mesurés totaux	Pourcentage d'élimination par Dixon
99-85-4	1-méthyl-4-propan-2-ylcyclohexa-1,4-diene	1	3	33
121-44-8	N,N-diethylethanamine	2	12	17
591-78-6	hexan-2-one	1	6	17
67-64-1	propan-2-one	2	12	17
109-99-9	oxolane (THF)	1	6	17
105-37-3	ethyl propanoate	1	7	14
123-86-4	butyl acetate	1	9	11
470-82-6	2,2,4-triméthyl-3-oxabicyclo[2.2.2]octane	1	9	11

Le pourcentage d'élimination par molécule varie entre 11% et 33%. En moyenne 17% des valeurs sont éliminées pour les 8 molécules considérées. Cela correspond à 10 valeurs sur un total de 321 mesures du référentiel, soit une élimination de 3% de l'ensemble de l'information. Cela souligne une stabilité de l'application du test de Dixon sur les mesures expérimentales et encourage à valider le caractère normal de la distribution des valeurs de seuils de perception par molécule. Cependant, le test de Dixon est une

méthode stricte qui n'élimine des données que si l'écart-type est déjà initialement relativement faible. Ainsi, le fait de ne pas éliminer beaucoup de valeurs peut également signifier que les mesures présentent un écart-type relativement large.

La répétabilité des mesures (Re) est calculée par molécule (Annexe 6). En raison de l'étendue de l'échelle des valeurs, comme évoqué au paragraphe 2.7.1.2, l'écart absolu moyen (EAM) individuel de chaque molécule, après standardisation, est calculé (Annexe 7). Ensuite, la moyenne et l'écart-type des écarts absolus moyens (EAM) individuels sont calculés. Ces données ont une moyenne de 0,8 pour un écart-type de 0,07. Cela signifie qu'en moyenne les observations de chacune des séries de molécules se trouvent à 0,8 unité de leur moyenne et cela avec une bonne précision puisque la distribution du référentiel est faible. Le référentiel est donc considéré comme précis.

2.7.2.2 Évaluation du nettoyage par molécule

Parmi les quarante-cinq molécules dont le seuil de perception a été déterminé expérimentalement, quatre ne sont pas renseignées dans la littérature. Sur les quarante et une valeurs de seuils de perception restantes, le nettoyage a mené à l'élimination des valeurs de la littérature pour huit molécules. Ainsi, la comparaison entre Z_1 et Z_2 est réalisée sur trente-trois molécules.

Afin de permettre la représentation de Z_1 , Z_2 et Re sur un même graphe la transformation indiquée en Tableau 18 est appliquée.

Tableau 18 : Définition des variables à comparer après transformation

Z'_1	$= \log_{10}(1,5 + Z_1)$
Z'_2	$= \log_{10}(1,5 + Z_2)$
Re'	$= \log_{10}(1,5 + Re)$

Additionner 1,5 permet de travailler dans la partie positive de la fonction logarithme décimal. Observer Z'_1 et Z'_2 permet de conclure directement sur le comportement de Z_1 et Z_2 (Annexe 8). Les trois variables du Tableau 18 sont représentées pour chacune des molécules sous forme d'histogramme pour visualiser sur un même graphique le nombre de fois où le nettoyage permet d'approcher le résultat expérimental (Figure 25).

L'arbitrage décrit dans le paragraphe 2.7.1.3 permet ensuite de juger de l'efficacité du nettoyage.

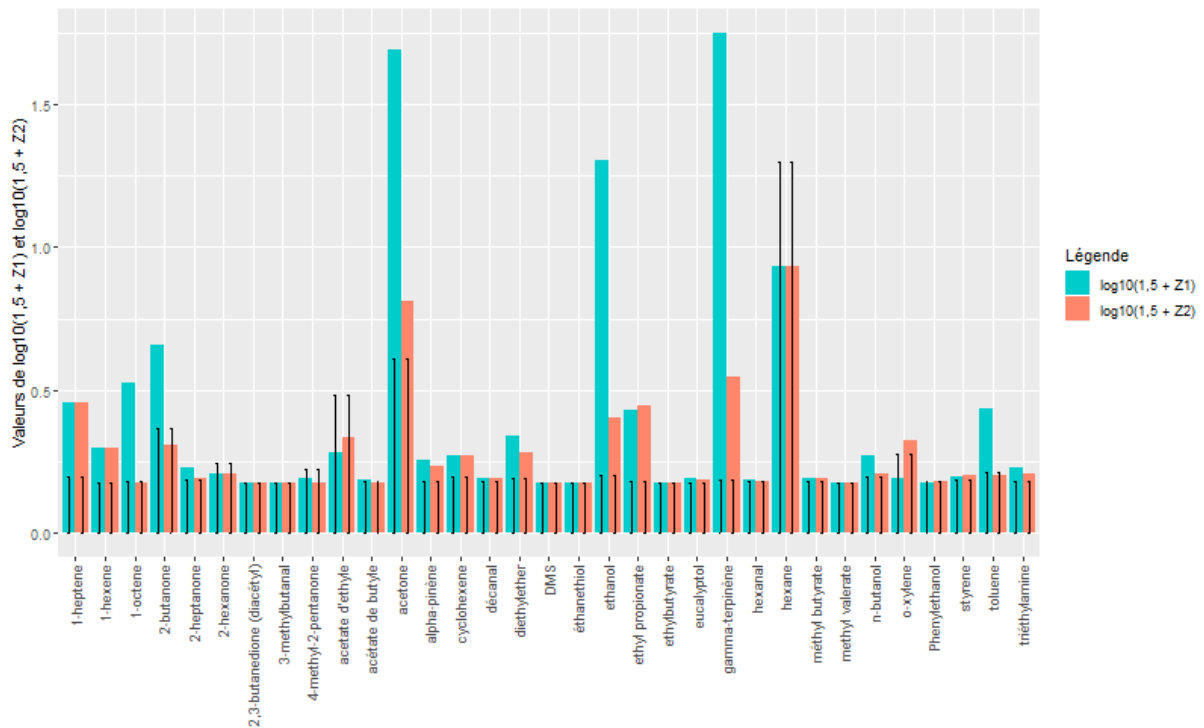


Figure 25 : Comparaison de Z'_1 , Z'_2 et Re' (en noir)

L'examen du graphe et des résultats de l'Annexe 8, fait ressortir que Z_2 est strictement inférieur à Z_1 dans dix-neuf cas sur trente-trois. Le nettoyage permet donc d'approcher les mesures expérimentales dans 60% des cas. Dans vingt-cinq cas sur trente-trois il y a infériorité ou égalité des deux valeurs (76% des cas). Il convient de souligner que si la transformation appliquée permet la représentation graphique conjointe de Z_1 , Z_2 et Re , elle atténue néanmoins la différence entre ces valeurs (Annexe 8).

Une comptabilisation des scénarios conduisant à une valeur « correcte » ou non, telle que définie au paragraphe 2.7.1.3, est présentée en Tableau 19.

Tableau 19 : Scénarios consécutifs au nettoyage

Scenario 1	Amélioration de la valeur même si valeur déjà « correcte »	1 cas
Scenario 2		... menant à une valeur « correcte »	6 cas
Scenario 3		... mais la valeur reste « incorrecte »	12 cas
Scenario 4	Pas d'amélioration de la valeur qui était déjà « correcte »	3 cas
Scenario 5		... qui reste « incorrecte »	3 cas
Scenario 6	Dégradation de la valeur mais qui reste « correcte »	2 cas
Scenario 7		... qui reste « incorrecte »	4 cas
Scenario 8		... qui devient « incorrecte »	2 cas

Avant nettoyage, huit valeurs de Z_1 sont « correctes » car inférieures à la répétabilité. Douze le sont après nettoyage.

Les scénarios 1, 2, 3, 4 et 6 sont considérés comme positifs car il s'agit des scénarios pour lesquels la valeur de Z_2 est « correcte » ou pour lesquels Z_2 est strictement inférieure à Z_1 ; ce qui correspond à

vingt-quatre cas positifs sur trente-trois. Ainsi, le nettoyage appliqué sur les données brutes de la littérature permet d'obtenir 73% de cas positifs sur les trente-trois cas analysés.

Les scénarios négatifs (5, 7 et 8) concernent neuf molécules dont les valeurs X_1, X_2, X_3, Z_1, Z_2 et de répétabilité (Re) sont présentées en Tableau 20.

Tableau 20 : Molécules correspondant aux trois scénarios négatifs (5,7, et 8)

N° CAS	Molécules	Re	X_1	X_2	X_3	Z_1	Z_2	Scénario
592-41-6	hex-1-ene	2,23E-03	4,80E-01	4,80E-01	5,22E-04	4,79E-01	4,79E-01	5
592-76-7	hept-1-ene	6,28E-02	1,50E+00	1,50E+00	1,43E-01	1,36E+00	1,36E+00	5
110-83-8	cyclohexene	7,44E-02	6,00E-01	6,00E-01	2,35E-01	3,65E-01	3,65E-01	5
623-42-7	methyl butanoate	1,14E-02	7,69E-02	7,70E-02	3,32E-02	4,37E-02	4,38E-02	7
431-03-8	butane-2,3-dione	4,21E-05	5,74E-04	1,93E-03	1,17E-04	4,56E-04	1,81E-03	7
105-37-3	ethyl propanoate	5,17E-03	1,20E+00	1,30E+00	2,22E-02	1,17E+00	1,28E+00	7
100-42-5	Styrene	2,34E-02	1,20E-01	1,42E-01	5,92E-02	6,03E-02	8,28E-02	7
60-12-8	2-phenylethanol	9,25E-03	1,44E-02	4,90E-03	1,53E-02	8,88E-04	1,04E-02	8
95-47-6	1,2-xylene	3,84E-01	1,44E+00	8,90E-01	1,49E+00	5,34E-02	6,03E-01	8

Le scénario 5 correspond à des molécules pour lesquelles aucune différence n'est constatée entre le seuil de perception de la littérature avant et après nettoyage mais qui correspond à une valeur supérieure à la répétabilité (Re). Ce scénario concerne trois des quatre alcènes du référentiel : hex-1-ene, hept-1-ene et cyclohexène. Pour ces trois molécules, le nettoyage n'améliore pas la donnée mais ne la dégrade pas non plus.

Le scénario 7 concerne quatre molécules. Parmi ces quatre molécules, les valeurs de Z_2 du méthyl butanoate et du styrène ne sont que très faiblement supérieures à la répétabilité. En revanche, l'éthyl propanoate et la 2,3-butanedione (butane-2,3-dione) présentent des répétabilités très faibles impliquant un intervalle de confiance très petit et les valeurs de Z_2 de ces deux molécules en sont fortement exclues. Pour ces deux molécules, il apparaît que le nettoyage n'est pas bénéfique.

Le scénario 8 concerne le 2-phenylethanol et le 1,2-xylene. La valeur de Z_2 de ces deux molécules est très proche de l'intervalle de confiance décrit par la répétabilité. L'efficacité globale du nettoyage peut, malgré tout, être validée pour ces deux molécules.

Cette analyse montre que quatre molécules sur les neuf correspondant à des scénarios défavorables sont néanmoins très proches de respecter les critères correspondant à un scénario favorable (méthyl butanoate, styrène, phényléthanol et o-xylène). Elles pourraient être incluses dans des scénarios positifs à la condition d'assouplir très légèrement les critères de sélection. Concernant les trois molécules du scénario 5, il ne s'agit pas de scénario strictement défavorable puisqu'il n'y a pas dégradation de la valeur. Sous cette réserve, seuls deux cas apparaissent comme strictement défavorables (l'éthyl propanoate et la 2,3-butanedione).

Pour compléter cette approche, un test statistique de significativité du nettoyage par molécule a été réalisé. Pour cela, cinq variables ont été définies par molécule (Tableau 21).

Tableau 21 : Variables définies par molécule

X_1 : Distribution des valeurs de seuils de perception bruts de la littérature ($\rightarrow N(m_1, \sigma_1)$)	m_1 : moyenne des valeurs de X_1 σ_1 : écart-type des valeurs de X_1
X_2 : Distribution des valeurs de seuils de perception après nettoyage ($\rightarrow N(m_2, \sigma_2)$)	m_2 : moyenne des valeurs de X_2 σ_2 : écart-type des valeurs de X_2
X_3 : Distribution des valeurs de seuils de perception expérimentaux ($\rightarrow N(m_3, \sigma_3)$)	m_3 : moyenne des valeurs de X_3 σ_3 : écart-type des valeurs de X_3
W_1 : $X_1 - X_3$ ($\rightarrow N(m_1 - m_3, \sqrt{\sigma_1^2 + \sigma_3^2})$)	
W_2 : $X_2 - X_3$ ($\rightarrow N(m_2 - m_3, \sqrt{\sigma_2^2 + \sigma_3^2})$)	

Un test de Student est ensuite appliqué pour détecter si W_1 et W_2 sont significativement différentes de zéro et si W_1 est significativement supérieure à W_2 par molécule. Cependant, en raison du faible nombre de données disponibles (Annexe 9), des données de seuils de perception très proches de zéro et des données de la littérature ayant un écart-type très large, les résultats obtenus se sont avérés très instables. Il a donc été choisi de réaliser un test statistique sur les moyennes de Z_1 et Z_2 . Un test de Fisher a permis de montrer que les deux variables présentaient des variances significativement différentes (p-value < $2,2 \times 10^{-16}$). L'utilisation du test de Welch, permet de constater que si les deux variables ne sont en moyenne pas significativement différentes avec un risque de 5% (p-value < 0,104), elles le sont en revanche si le niveau de risque est porté à 10%.

Enfin, pour vérifier la pertinence de l'élimination des seuils de perception des huit molécules à l'étape 2, les valeurs de Z_1 de ces molécules ont été comparées à la répétabilité (Re) (Tableau 22).

Tableau 22 : Molécules dont toutes les valeurs de seuils de perception ont été éliminées par le nettoyage

N° CAS	Molécules	Re	X_1	X_2	X_3	Z_1	Z_2
120-92-3	cyclopentanone	0,271	31	NC	0,910	30,089	NC
5989-27-5	(4R)-1-methyl-4-prop-1-en-2-ylcyclohexene (R-Limonène)	0,023	0,152	NC	0,082	0,069	NC
104-76-7	2-ethylhexan-1-ol	0,266	0,799	NC	0,380	0,418	NC
140-11-4	benzyl acetate	0,168	85	NC	0,172	84,827	NC
1634-04-4	2-methoxy-2-methylpropane (MTBE)	0,012	0,628	NC	0,024	0,604	NC
98-86-2	1-phenylethanone	0,003	0,139	NC	0,012	0,126	NC
109-99-9	oxolane (THF)	0,550	89,859	NC	2,269	87,589	NC
108-83-8	2,6-dimethylheptan-4-one	0,199	9,282	NC	0,400	8,882	NC

Ce tableau permet de constater que toutes les valeurs de Z_1 sont supérieures à la répétabilité. Cela permet de souligner l'intérêt du nettoyage qui a permis d'éliminer des valeurs en dehors de l'intervalle de confiance même si son application peut conduire à ne conserver aucune valeur de seuil.

2.8 Conclusion sur le protocole de nettoyage

Un protocole en cinq étapes permettant de fiabiliser les données de seuils de perception de la littérature a été proposé.

La première étape a consisté à constituer une base de données recensant, pour 650 molécules, des données de seuils de perception répartis sur 165 articles.

À partir de cette base, une seconde étape a consisté à éliminer les valeurs issues des articles les moins fournis en nombre de seuils de perception, donc considérés comme *a priori* les moins fiables. Cela a conduit à l'élimination des articles contenant strictement moins de quatre mesures de seuils de perception. Cela correspond à moins de 10% du total des données de la base. Lors de cette étape, plus de la moitié des articles a été éliminée tout en conservant 97 % des molécules renseignées par au moins une valeur.

À l'étape 3, pour chaque molécule, les valeurs aberrantes ont été éliminées à l'aide du test de Grubbs pour un risque à 5%. Cette étape a conduit à éliminer 10% du total des données de la base.

À l'étape 4, un taux de comparaison et une fréquence d'exclusion de chaque article ont été définis. Seuls les articles ayant successivement un taux de comparaison strictement supérieur à 40% puis une fréquence d'exclusion inférieure ou égale à 12% ont été conservés. Cette dernière limite a été fixée à partir d'une classification ascendante hiérarchique. Cette étape a conduit à éliminer environ 50% de l'ensemble des données. Accepter un taux de perte d'information aussi élevé est justifié par la volonté de privilégier la fiabilité des données plutôt que leur nombre.

Un bilan global du protocole de nettoyage des seuils de perception appliqué à la base de données est présenté dans le Tableau 23.

Tableau 23 : Bilan de la perte d'information totale liée au nettoyage

	Ensemble du nettoyage		
	Avant	Après	Pourcentage de PI
Nombre d'articles	165	33	80%
Nombre de seuils de perception	1503	614	59,1%
Nombre de molécules renseignées par au moins une valeur de seuil de perception	650	338	48%

Moins de la moitié des molécules renseignées par au moins une valeur de seuil de perception est éliminée. Le nombre de molécules restantes est de 338. Ce nombre reste très raisonnable pour envisager le travail de complétude des données de seuils de perception à partir de l'existant, en chapitre 3.

La cinquième étape a été l'évaluation du nettoyage. Pour cela, les valeurs de seuils de perception de la littérature conservées ont été comparées aux trente-trois seuils de perception mesurés constituant notre référentiel. Dans vingt-huit cas, il est constaté une évolution favorable du seuil de perception (85%) ; dans trois cas, le nettoyage n'est ni favorable ni défavorable (9%) et dans deux cas, le scénario est strictement défavorable (6%). Un test de significativité statistique a montré qu'en prenant un risque de 5%, le nettoyage ne permet pas d'approcher en moyenne la valeur expérimentale. Il le permet cependant si l'on accepte un risque de 10%.

Enfin, il convient de souligner que l'application de la démarche globale de nettoyage des seuils de perception sur la base des données initiales, en y intégrant les mesures réalisées par Olentica sous la forme d'un article, a abouti à la conservation de l'ensemble des seuils mesurés attestant, là encore, de la pertinence du nettoyage.

CHAPITRE 3 : AMELIORATION DU DEGRE DE COMPLETEUDE DES DONNEES DE SEUILS DE PERCEPTION

3.1 Contexte

Comme vu dans le chapitre précédent, le nombre de molécules possédant une valeur de seuil de perception renseignée dans la littérature est très faible au regard du nombre de molécules existantes (à minima 96 540 411 molécules recensées dans PubChem (NIH 2004)). Le nettoyage de la base de données disponible présenté au chapitre 2 a abouti à une réduction du nombre de molécules renseignées de 650 à 338. Or, pour identifier les molécules responsables de l'odeur, il est indispensable de disposer d'une valeur de seuil de perception pour toutes les molécules mises en évidence lors des analyses.

Deux solutions sont alors possibles :

- La mesure du seuil de perception comme présenté dans le paragraphe 2.7.1.1. Cette solution est inenvisageable au vu du très grand nombre de molécules existantes non renseignées.
- Le développement d'un modèle de prédiction des seuils de perception inconnus sur la base des seuils de perception pouvant être considérés comme les plus robustes (les 45 seuils de perception mesurés au laboratoire auxquels sont ajoutés les seuils de perception de la littérature retenus suite à la procédure de nettoyage présentée au chapitre 2). La base de données ainsi constituée est composée de 350 molécules renseignées.

La construction d'un modèle pertinent et stable passe par plusieurs étapes :

- L'identification d'une méthode de modélisation prédictive adaptée aux données et aux objectifs du projet ;
- La sélection des propriétés physico-chimiques pouvant être reliées au seuil de perception olfactive des molécules considérées. Il est indispensable que les grandeurs physiques liées à ces propriétés soient disponibles pour toutes les molécules présentes dans la base de données. Ces données constitueront les variables d'entrée du système ;
- La division de la base de données en deux jeux de données distincts couvrant au mieux la diversité des structures chimiques rencontrées et celle des valeurs de seuils de perception. Seront ainsi constitués :
 - Un jeu de données qui servira à la construction du modèle (« train ») ;
 - Un jeu qui servira à l'évaluer (« test ») ;
- Le choix de la fonction mathématique permettant la corrélation des données ;
- L'affinage du modèle par l'analyse des écarts de prédiction des seuils de perception.

Ces cinq étapes sont détaillées dans la suite de ce chapitre.

3.2 Choix de l'approche du modèle

Cette partie consiste à exposer le type d'approche sur lequel nous sommes orientés pour ce projet.

3.2.1 Méthodologie

Les variables d'entrée du modèle doivent être des grandeurs pouvant être reliées aux propriétés odorantes des molécules. Ces grandeurs sont (MEIERHENRICH et al. 2005) :

- La masse molaire qui peut être un indicateur de l'encombrement stérique d'une molécule ;
- La polarité ;
- La pression de vapeur saturante ;
- Le caractère lipophile qui conditionne l'interaction entre les molécules volatiles et la muqueuse olfactive.

Le caractère lipophile d'une molécule est représenté par le coefficient de partage octanol-eau (pK_{ow}) (AIZPURU 2001). Cette grandeur est disponible dans la banque de données PubChem (NIH 2004) sous la forme d'un indice (XlogP3) obtenu par modélisation [(WANG 2007) ; (CHENG et al. 2007)].

La masse molaire et la pression de vapeur saturante sont, elles, disponibles dans des banques de données telles que la Base NIST (NIST 2017), le site PubChem (NIH 2004) ou le site ChemSpider (ROYAL SOCIETY OF CHEMISTRY 2015).

Si l'on ne considère que ces trois grandeurs, il apparaît que si la masse molaire et le coefficient de partage octanol-eau sont disponibles pour toutes les molécules de la base, les données concernant la pression de vapeur saturante sont, elles, peu disponibles.

Ce manque de disponibilité des données pour l'ensemble des molécules nous a incités à aborder la problématique par le biais d'une approche globale : l'approche RQSA/RQSP (plus connue sous les termes QSAR/QSPR pour « Quantitative Structure-Activity Relationship/Quantitative Structure-Property Relationship »). Cette approche consiste à prédire une grandeur physique (activité, propriété) d'une molécule à l'aide d'indices traduisant la structure de cette molécule (ROY, KAR, et DAS 2015). Elle permet d'envisager la prédiction des seuils de perception manquants à partir d'une base de données complète composée d'un nombre important d'indices de structure disponibles pour toutes les molécules.

3.2.2 Les modèles de type RQSA/RQSP et leur mise en œuvre

Les modèles RQSA/RQSP sont utilisés pour corréler les propriétés physiques, chimiques, biologiques ou sensorielles de composés à leurs caractéristiques structurales (POLSTER et SCHIEBERLE 2015).

Dans les années soixante, des chercheurs se sont intéressés à la topographie des molécules, l'idée directrice de leurs recherches étant que cette topographie influençait les propriétés de la molécule. Par exemple, Hansch (1979) a défini le coefficient de partage octanol-eau (K_{ow}) sur la base des propriétés d'hydrophobicité et électroniques des molécules (HANSCH et LEO 1979). À partir de cette époque, de

nombreux indices topologiques calculés à partir d'informations structurales, et codant des données liées aux propriétés physico-chimiques des molécules ont alors été proposés. On en compte aujourd'hui plus de 6000 (TODESCHINI et CONSONNI 2009b).

La mise en place d'un modèle RQSA/RQSP se fait en deux étapes. La première consiste à déterminer les paramètres structuraux descriptifs de la propriété (ou indices). La seconde à établir le modèle lui-même à l'aide d'outils mathématiques.

Les indices pris en compte pour décrire la propriété sont généralement définies de manière plus ou moins intuitive. Si un grand nombre d'indices est retenu, il est possible de réduire leur nombre en excluant les paramètres fortement corrélés avec d'autres, et ceux ayant une faible variance, donc contribuant peu ou pas à l'explication de la variance de la propriété.

Le modèle mathématique permettant d'établir la relation entre la propriété considérée et les paramètres structuraux de la molécule revient à trouver la fonction f telle que $y = f(x_1, x_2, \dots, x_n)$, où x_1, x_2, \dots , et x_n sont les différents termes structuraux descriptifs. Même si elle peut prendre les formes les plus diverses, la fonction f est généralement de type linéaire.

En d'autres termes, établir un modèle RQSA/RQSP consiste essentiellement en la recherche analytique d'un modèle reliant la propriété modélisée à des variables indépendantes. Deux points de vigilance sont à souligner : le nombre de paramètres à prendre en compte et la possibilité de corrélations simplement dues au hasard. Les relations développées sont donc toujours évaluées en termes statistiques.

Comme mentionné plus haut, cette approche est fréquemment utilisée en physique, chimie et biologie. Par exemple, des études ont concerné la puissance répulsive d'une molécule contre les moustiques [(MA et al. 1999) ; (NATARAJAN et al. 2005) ; (BASAK et al. 2007)], les moments dipolaires des molécules hydrocarbonés (NESTEROV et al. 2004), l'acidité des cétones (YUAN, MOSIER, et ZHANG 2012), la solubilité dans l'eau d'un nombre importante de molécules (TOROPOV, TOROPOVA, BENFENATI, et al. 2013), l'absorption de nanoparticules dans les cellules cancéreuses (TOROPOV, TOROPOVA, PUZYN, et al. 2013), ou encore la prédiction de la température d'ébullition d'alcane (RAWAT et SATI 2013). Cette approche est aussi utilisée pour l'aide à la création de nouveau médicaments (GOULON-SIGWALT-ABRAM 2008).

Ces différentes études ont fourni des résultats encourageants quant à l'efficacité de l'approche RQSA/RQSP à l'aide de traitements statistiques linéaires.

Certains auteurs se sont penchés sur la problématique de la mise en relation de la structure des molécules avec l'odeur et les seuils de perception [(DEMOLE et STOLL 1964) ; (ABRAHAM et al. 2002) ; (CZERNY et al. 2011) ; (STRUBE, BUETTNER, et CZERNY 2012) ; (JUHLKE et al. 2017)]. D'autres

ont tenté de réaliser un modèle prédictif des seuils de perception à partir d'une approche QSAR/QSPR [(X. XU et al. 2012) ; (ZARZO 2012) ; (PAL, MITRA, et ROY 2014) ; (POLSTER et SCHIEBERLE 2015) ; (TOROPOV et al. 2016)] dont les informations essentielles sont présentées dans le Tableau 24. Cette démarche est justifiée par le fait que l'encodage des odeurs réside dans un système de reconnaissance substrat/récepteur complexe (DE MARCH 2015) et que la complexité moléculaire déterminerait le nombre de notes olfactives et l'acceptabilité d'une odeur (KERMEN et al. 2011).

Tableau 24 : Résumé d'études statistiques portant sur la prédiction du seuil de perception de molécules (RN = Réseau de Neurons ; RL= régression Linéaire ; RLM = Régression Linéaire Multiple ; PLS = « Partial Least square Regression » ; SVM=Support Vecteur Machine ; »

Articles	Source SP	Indices RQSA/RQSP	Outil mathématique utilisé	Types et nombre de molécules analysées
(X. XU et al. 2012)	(LI, YU, et LI 1992)	Oui	RLM et RN et SVM	50 hétérocycles contenant des atomes d'oxygène et d'azote
(ZARZO 2012)	(Devos et al., 1990)	Oui (longueur de la chaîne de carbone)	RLM et non RLM (StatGraphics)	114 molécules aliphatiques
(PAL, MITRA, et ROY 2014)	(Anker et al., 1990)	Oui	GFA-RLM et G-PLS	53 alcools aliphatiques
(POLSTER et SCHIEBERLE 2015)	NC	Oui	NC	46 alcanes thiols
(TOROPOV et al. 2016)	(Van Gemert, 2011)	Oui avec méthode Monte Carlo	RL	906 molécules diverses

À l'exception du modèle proposé par Toropov et al. (TOROPOV et al. 2016), il s'agit de modèles locaux construits sur un faible nombre de molécules ayant une spécificité structurale (Tableau 24). Ces différentes études ont fourni des résultats encourageants quant à l'efficacité de l'approche QSAR/QSPR à l'aide de traitements statistiques variés, aussi bien linéaires que non linéaires (réseaux de neurones, support vecteur machine). Cependant, les modèles construits sont satisfaisants pour les données analysées mais dès lors que les molécules n'appartiennent pas à l'espace de construction, la fiabilité de la réponse devient plus aléatoire. Cela ne permet pas de répondre au besoin de disposer d'une valeur de seuil de perception quelle que soit la molécule rencontrée.

Seul Toropov (TOROPOV et al. 2016) a construit un modèle de prédiction des seuils olfactifs à partir d'un jeu de données de molécules variées et de taille importante.

Un apport important de notre projet par rapport à ces travaux, hors ceux de Toropov (TOROPOV et al. 2016), réside dans le fait qu'il vise à construire un modèle général pour un jeu de données de taille importante basé sur un ensemble de molécules de natures chimiques diverses. L'une des différences de notre approche par rapport à celle de Toropov est de définir spécifiquement les données de construction du modèle pour une plus grande stabilité.

La modélisation du seuil de perception olfactive à partir d'une approche RQSA/RQSP apparaissant envisageable au travers des quelques exemples donnés plus haut, il est maintenant nécessaire de sélectionner les indices qui vont servir de variables d'entrée au modèle.

3.3 Choix des indices

Il existe un nombre important d'indices permettant de décrire la structure des molécules selon les différents plans de l'espace. Todeschini et Consonni ont inventorié 6400 indices parus sur la période 1741 – 2008 (TODESCHINI et CONSONNI 2009b). Il est donc avant tout nécessaire de sélectionner les indices les plus pertinents pour notre étude.

3.3.1 Critère de sélection des indices

Cette sélection des indices se fait selon deux critères :

- Les indices pouvant avoir un impact sur les propriétés odorantes des molécules (encombrement stérique, caractère lipophile, capacité d'interaction des molécules).
- La possibilité de calculer ces indices sélectionnés pour l'ensemble des molécules.

Plusieurs logiciels permettent le calcul d'indices de structure :

- « Dragon » [(CONSONNI et al. 2002) ; (SOUICI, LOURICI, et MESSADI 2007) ; (HEMMATEENEJAD, MIRI, et ELYASI 2012) ; (GHARAGHEIZI et al. 2012) ; (TOROPOVA et al. 2015)] capable de calculer 3000 descripteurs (GHARAGHEIZI et al. 2012) ;
- « HyperChem » [(MALDONADO 2006) ; (SOUICI, LOURICI, et MESSADI 2007) ; (GHARAGHEIZI et al. 2012) ; (HEMMATEENEJAD, MIRI, et ELYASI 2012)] ;
- « Cerius2 » pour des indices de différentes nature (PAL, MITRA, et ROY 2014) ;
- « PaDEL » (PAL, MITRA, et ROY 2014) pour des indices topologiques ;
- ACD/ChemSketch 5.12 [(MALDONADO 2006) ; (TOROPOVA et al. 2015)] pour la polarisabilité.

Il arrive que plusieurs de ces logiciels soient utilisés en complémentarité dans une même étude (PAL, MITRA, et ROY 2014). En effet, ces logiciels ne couvrent parfois pas tous les types de molécules et les indices calculés ne sont pas toujours suffisants. De plus, dans ces logiciels, les indices sont calculés à partir d'une écriture simplifiée mais non universelle des molécules : l'écriture SMILES (« Simplified Molecular Input Line Entry Specification ») [(WEININGER 1988) ; (O'DONNELL 2009)]. Ce manque d'unicité rend difficile la compatibilité entre logiciels puisque chacun d'eux n'utilise pas obligatoirement la même forme de codage du SMILES (O'DONNELL 2009). L'absence de mises à jour régulières, les nombreux dysfonctionnements et la contrainte liée à la licence payante rendent complexe l'utilisation de ces logiciels (MORIWAKI, TIAN, et TAKAGI 2018).

C'est pour cette raison que nous avons choisi de calculer l'ensemble des indices d'intérêt à l'aide du logiciel libre R. Cette approche est, dans l'esprit, similaire à celle d'autres auteurs qui ont mis en place un logiciel de calcul d'indices sous Python (MORIWAKI, TIAN, et TAKAGI 2018).

Le calcul des indices nécessaires est effectué à partir de deux identifiants disponibles pour l'intégralité des molécules : il s'agit de l'InChI (« International Chemical Identifier ») et du SMILES [(WEININGER 1988) ; (O'DONNELL 2009)]. Contrairement au SMILES, l'InChI est un identifiant universel de la molécule. La source internet utilisée pour extraire l'InChI est la base NIST (NIST 2017) et celle utilisée pour extraire le SMILES est principalement la base PubChem (NIH 2004). Un exemple de ces représentations est donné en Figure 26 pour la molécule d'indole. L'identifiant SMILES est utilisé sous la forme « Kekulé » (O'DONNELL 2009).

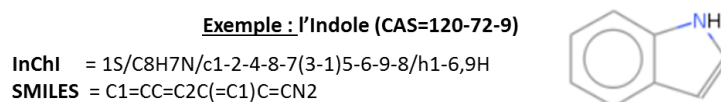


Figure 26 : Illustration de l'InChI et du SMILES pour l'indole

La fonction construite sous R permet de calculer 69 indices de structure d'intérêt dans la problématique des odeurs. Ils peuvent être répertoriés en cinq catégories :

- **Indice de composition:** Nombre et nature des atomes, nombre et nature des liaisons, nombre et nature des carbones, nombre et nature des groupes fonctionnels ;
- **Indices de constitution:** Nombre de cycles, masse molaire, moyenne des masses molaires individuelles, nombre d'insaturations ;
- **Indices topologiques** (dans l'ordre chronologique d'apparition): Wiener, Zagreb, Randic et connectivité moléculaire, Balaban, Schultz, Harary, Gutman ;
- **Indices électrotopologiques:** E-state, ZEP, électronégativité de Sanderson ;
- **Indice de lipophilie :** XlogP3

L'identification et la comptabilisation des groupes fonctionnels est rendue nécessaire par le fait que les valeurs de seuils de perception sont généralement liées à la présence ou à l'absence de ces groupes dans les molécules[(DEMOLE et STOLL 1964) ; (ZARZO 2012)].

Les indices calculés et leur notation sont présentés dans le Tableau 25. Les calculs et les sources bibliographiques utilisées sont présentés en Annexe 10.

Tableau 25 : Liste des variables d'entrée et leur notation

Indices de composition		Indices topologiques	
Codage	Variables d'entrée	Codage	Variables d'entrée
C	Nombre d'atomes de carbone	W.Dh.	Indice de Wiener appliqué sur Dh
H	Nombre d'atomes d'hydrogène	M1	Indice de Zagreb M1
Br	Nombre d'atomes de brome	M2	Indice de Zagreb M2
Cl	Nombre d'atomes de chlore	X1Xp	Indice de conductivité moléculaire
F	Nombre d'atomes de fluor	X0X	Indice de conductivité moléculaire
I	Nombre d'atomes d'iode	X1X	Indice de conductivité moléculaire
N	Nombre d'atomes d'azote	X0Xv	Indice de conductivité moléculaire
O	Nombre d'atomes d'oxygène	X1Xv	Indice de conductivité moléculaire
P	Nombre d'atomes de phosphore	J.Dh.	Indice de Balaban appliqué sur Dh
S	Nombre d'atomes de soufre	Jt.Dh.	Indice de Balaban appliqué sur Dh
Si	Nombre d'atomes de silice	MTI	Indice de Schultz appliqué sur Dh
NbLtot	Nombre de liaisons total	HararyDh	Indice de Harary appliqué sur Dh
NbLS	Nombre de liaisons simples	GUTMAN	Indice de GUTMAN
NbLD	Nombre de liaisons doubles	W..Ah.	Indice de Wiener appliqué sur Ah
NbLT	Nombre de liaisons triples	J..Ah.	Indice de Balaban appliqué sur Ah
NbLConj	Nombre de liaisons conjuguées	Jt..Ah.	Indice de Balaban appliqué sur Ah
nCprimaire	Nombre de carbones primaires	Indices de constitution	
nCsecondaire	Nombre de carbones secondaires	Codage	Variables d'entrée
nCtertiaire	Nombre de carbones tertiaires	MM_Molecule	Masse molaire
nCquaternaire	Nombre de carbones quaternaires	AMW	Moyenne des masses molaires individuelles
Acides	Nombre de fonctions acides	Ni	Nombre d'insaturation
Alcane	Présence ou non d'un alcane	CYCLE	Nombre de cycles
Alcène	Nombre de double liaison "alcènes"	Indice de lipophilie	
Alcool	Nombre de fonctions alcools	Codage	Variables d'entrée
Alcynes	Nombre de triple liaison "alcynes"	XLogP3	Coefficient de partage octanol-eau
Aldéhydes	Nombre de fonctions aldéhydes	Indices électrotopologiques	
Amides	Nombre de fonctions amides	Codage	Variables d'entrée
Amines	Nombre de fonctions amines	StTot	Indice E-State
Anhydrides	Nombre de fonctions anhydrides	ZEP	indice Berinde basé sur CEP
Aromatiques	Nombre de cycles aromatiques	Electronegativity	Indice de l'électronégativité de Sanderson
Bromes	Nombre de fonctions bromées	W...CEP.	Indice de Wiener appliqué sur CEP
Cétones	Nombre de fonction cétones	J...CEP.	Indice de Balaban appliqué sur CEP
Chlores	Nombre de fonctions chlorées	Jt...CEP.	Indice de Balaban appliqué sur CEP
Esters	Nombre de fonctions esters		
Ethers.	Nombre de fonctions éthers		
Fluores	Nombre de fonctions fluores		
Iodes	Nombre de fonctions iodées		
Phosphores	Nombre de fonctions phosphores		
Soufres	Nombre de fonctions soufrées		
Terpènes	Présence ou non d'un terpène		
Peroxydes	Nombre de fonctions peroxydes		
Sulfinyles	Nombre de fonctions sulfinyles		

L'ensemble de ces indices sélectionnés pour cette étude est calculable pour toutes les molécules ayant un identifiant InChI et SMILES à l'exception des composés silylés.

3.3.2 Résultat : Base de données d'entrée

Une base de données complète de soixante-neuf indices est disponible pour les trois cent cinquante molécules dont la valeur du seuil de perception est validée. Cette nouvelle base de données (BDD2) est présentée en Annexe 11 et Annexe 12.

Ces soixante-neuf indices servent de variables d'entrée pour la prédiction des seuils de perception manquants. Une variable secondaire peut être formée par combinaison des indices de présence des groupes fonctionnels individuels. Cela permet de tracer une carte de chaleur (HeatMap) croisant la famille chimique attribuée à chacune des molécules avec les groupes fonctionnels présents dans la molécule (ou sous-famille chimique) de la base de données (Figure 27).

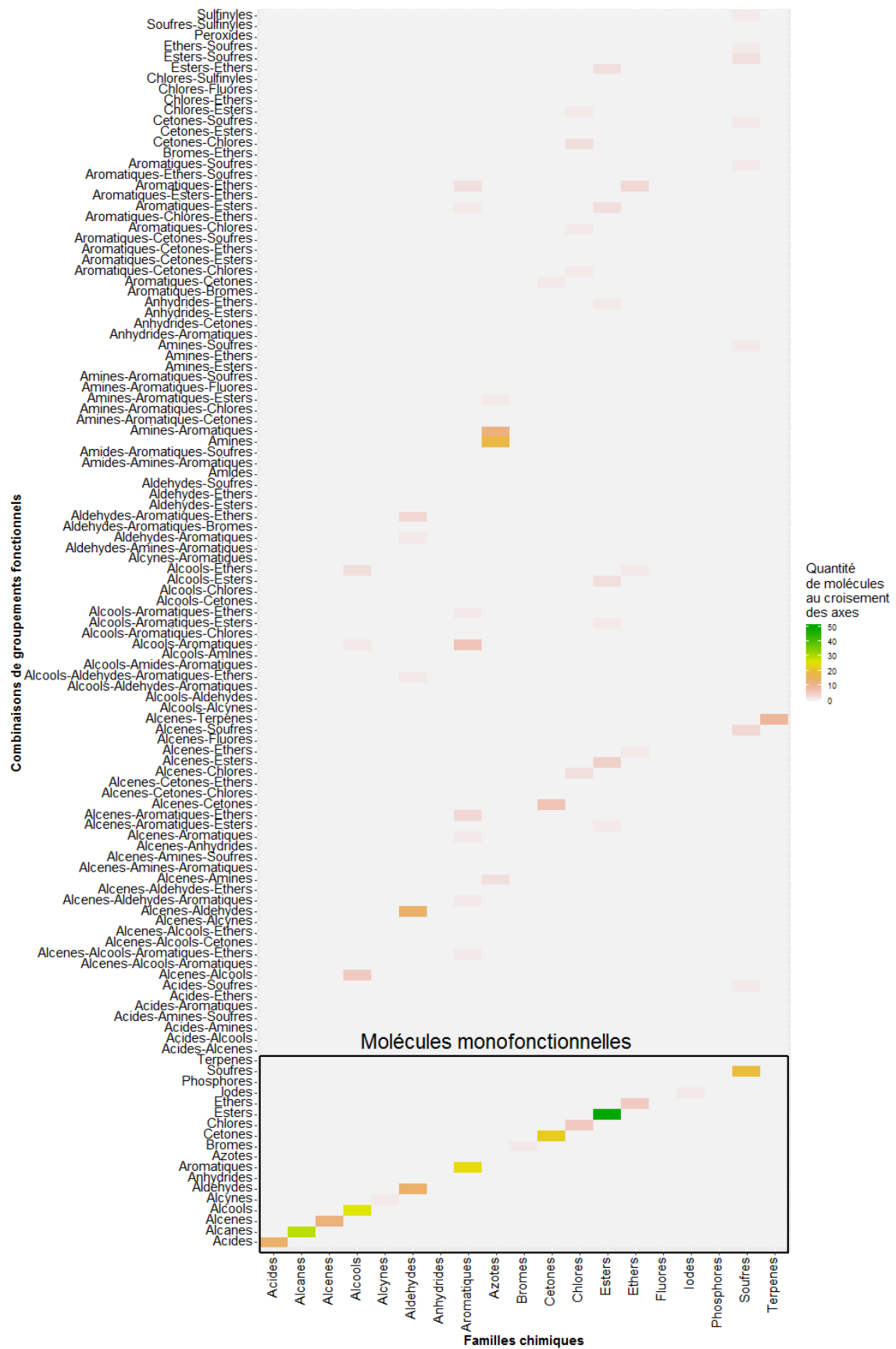


Figure 27 : HeatMap de la répartition des familles chimiques en fonction des combinaisons de groupements fonctionnels

Cette figure permet de confronter les deux types de regroupement des 350 molécules et de constater que 64% des molécules de la base sont monofonctionnelles. Toutes les autres contiennent au moins deux groupements fonctionnels différents et leur appartenance à l'une ou l'autre des familles chimiques peut être discutée. Étant donnée la diversité des molécules inventoriées, cela souligne la variabilité liée au choix de l'expert lors de la sélection d'un seuil de perception pour une molécule non-renseignée.

3.4 Sélection du jeu de construction

Ce paragraphe propose une méthodologie de division du jeu de données maîtrisée et contrôlée en données de construction (« train ») et en données de validation (« test ») en vue de générer le jeu de construction du modèle de prédiction des seuils de perception.

3.4.1 Méthodologie

L'étape primordiale pour la construction du modèle est la sélection du jeu de données d'apprentissage et le jeu de données de validation sur les trois cent cinquante molécules de la base de données.

Le jeu de données d'apprentissage doit être le plus représentatif possible de la variabilité des ordres de grandeur des seuils de perception et des différentes structures rencontrées dans la base de données. Dans cette optique, il y a nécessité de catégoriser les molécules selon leur structure. Une catégorisation stable passe par une réduction du nombre d'indices pour éliminer le bruit, suivie d'une synthèse de l'ensemble des indices sur les projections sous-jacentes les plus intéressantes.

Plus précisément, la réduction du bruit consiste à éliminer la redondance et la colinéarité des indices. Le traitement utilisé pour cette sélection est la méthode Lasso (WAINWRIGHT 2009) avec un calcul d'optimisation proximale (BLONDEL, SEKI, et UEHARA 2013). Les soixante-neuf indices étant de natures différentes (réelle, booléenne ou entière), leur sélection par la méthode Lasso est réalisée par type de données. À cette fin, un traitement spécifique est appliqué à chaque type de données :

- Une standardisation centrée réduite pour les indices réels ;
- Pas de standardisation pour les indices booléens ;
- Une standardisation dans $[-1,1]$ pour les indices entiers.

Les indices conservés par la méthode Lasso sont ensuite réassemblés pour déterminer les projections les plus intéressantes.

Les deux techniques les plus courantes permettant de trouver ces projections sont l'Analyse en Composantes Principales (ACP) (JOLLIFFE 2002) et la « Projection Pursuit » (PP) [(FRIEDMAN et TUKEY 1974) ; (FRIEDMAN 1985) ; (HUBER 1985) ; (JONES et SIBSON 1987)]. La principale différence entre ces deux approches est la manière dont les projections, ou composantes, sont construites.

- Dans le cas de l'Analyse en Composantes Principales, les composantes sont construites de manière à maximiser la variance globale du jeu de données. Cette technique ne permet donc pas

de distinguer les distributions gaussiennes des distributions non gaussiennes (BACHMANN et DONATO 2000).

- A contrario, la « Projection Pursuit » permet de projeter la distribution la moins gaussienne possible et permet ainsi de trouver des projections asymétriques ou multimodales afin de révéler une structure de classes sous-jacente (BACHMANN et DONATO 2000). Autrement dit, la « Projection Pursuit » permet de maximiser l'entropie, c'est-à-dire le désordre.

L'Analyse en Composantes Principales est généralement décrite comme moins robuste que la « Projection Pursuit » [(XIE et al. 1993) ; (BACHMANN et DONATO 2000)]. En se basant sur les recherches de Huber (HUBER 1985) et Jones et Sibson (JONES et SIBSON 1987), Hyvarinen et Oja soutiennent que la distribution gaussienne est la moins intéressante [(FRIEDMAN 1985) ; (HYVARINEN et OJA 2000) ; (CHOUAF et SMARA 2012)]. Enfin, Bachmann et Donato montrent que la « Projection Pursuit » permet d'améliorer la séparation des données sur une application géologique (BACHMANN et DONATO 2000).

Ces raisons expliquent pourquoi la « Projection Pursuit » est préférée à l'Analyse en Composantes Principales pour séparer les données suivant leurs divergences. L'algorithme appliqué est présenté dans le package « fastICA » du logiciel R (MARCHINI et al., 2017a). Il s'agit d'un package construit à partir de l'article de Hyvärinen (HYVARINEN et OJA 2000). Dans cette approche, le critère maximisé par composante est la néguentropie, ou entropie négative, qui est un facteur caractérisant le degré d'ordre d'un système et s'opposant à la tendance naturelle à la désorganisation (LAROUSSE 2018). La néguentropie est un critère considéré comme plus robuste et plus rapide à obtenir que l'indice de Kurtosis [(LENNON et al. 2001) ; (MARCHINI, HEATON, et RIPLEY 2017)]. Celui-ci est, tout comme la néguentropie, une mesure du caractère gaussien d'une variable : le caractère gaussien d'une variable diminuant avec une augmentation de la néguentropie ou de l'indice de Kurtosis (HYVARINEN et OJA 2000). Ainsi, plus une structure est néguentropique, plus elle maximise le désordre autour d'elle [(RODDIER 2012) ; (STAUNE 2015) ; (JAVEAU 2016)]. Cela implique que la maximisation de la néguentropie pour construire les projections sous-jacentes des indices permet donc de maximiser le désordre de ceux-ci.

Dans l'approche « Projection Pursuit », les composantes de projection sont choisies de manière séquentielle de façon à maximiser l'entropie à chaque séquence. Cet aspect séquentiel de construction des projections implique un ordre de priorité des axes (LENNON et al. 2001). Cela implique que l'axe 1 (ou direction 1) suffit pour représenter au mieux le caractère non gaussien de la distribution des données dans cette direction [(ROBERTS et EVERSON 2001) ; (CHOUAF et SMARA 2012)] et que l'observation des différences entre molécules est priorisée sur cet axe 1 bien que chaque axe ait son importance. Pour ces raisons et pour une praticité de traitement, nos données sont visualisées par la suite selon les deux premiers axes (ou projections).

L'application de la « Projection Pursuit » permet de récupérer la matrice des coordonnées des molécules sur les projections 1 et 2 (matrice notée S). Cette matrice S est ensuite utilisée pour catégoriser les molécules (ROBERTS et EVERSON 2001).

Cette catégorisation est réalisée par une approche de Classification Ascendante Hiérarchique (CAH), consolidée itérativement par l'indice de Hopkins. L'indice de Hopkins est un indicateur introduit par Hopkins et Skellam (HOPKINS et SKELLAM 1954) permettant d'indiquer si un jeu de données tend vers une structure classable par comparaison à un jeu de données simulé aléatoirement selon une distribution uniforme (LAWSON et JURIS 1990). Il permet donc de faire la nuance inter et intra classes. La valeur de l'indice de Hopkins (H) varie principalement entre 0,5 signifiant une structure aléatoire uniforme et 1 correspondant à un jeu de données contenant des classes [(LAWSON et JURIS 1990) ; (RUNKLER 2012) ; (GUIGOURES 2013) ; (LARSEN 2014)]. Plus précisément, une valeur d'indice de Hopkins supérieure à 0,75 permet d'établir avec un niveau de confiance de 90% que le jeu de données a une tendance à la classification [(LAWSON et JURIS 1990) ; (DE MAESSCHALCK et al. 1999)]. Il peut arriver qu'une valeur d'indice de Hopkins s'approche de 0 correspondant alors à une distribution des points selon une grille régulière (RUNKLER 2012). Nous définissons, pour notre étude, un indice de Hopkins minimum de 0,75 pour déterminer le caractère classable de nos données.

Le calcul de l'indice de Hopkins nécessite de sélectionner un sous espace de grandeur notée « n ». Or, comme l'évoque Adolfsson, le choix optimal de « n » doit permettre de sélectionner un espace représentant entre 5% et 10% du jeu de données total (ADOLFSSON, ACKERMAN, et BROWNSTEIN 2017). Dans notre cas, « n » est choisi de manière à sélectionner un espace représentant 10% du jeu de données. Ce sous-échantillonnage est réalisé 30 fois et la valeur finale de l'indice d'Hopkins (H) correspond au mode² maximal de la série des 30 valeurs obtenue suite aux 30 itérations. La fonction utilisée pour le calcul d'une valeur d'Hopkins est décrite en Annexe 13.

Cette démarche de consolidation de la Classification Ascendante Hiérarchique par l'indice de Hopkins en combinaison simultanée est présentée en Figure 28.

² Valeur d'un caractère quantitatif discret ayant la fréquence la plus élevée (LAROUSSE 2018)

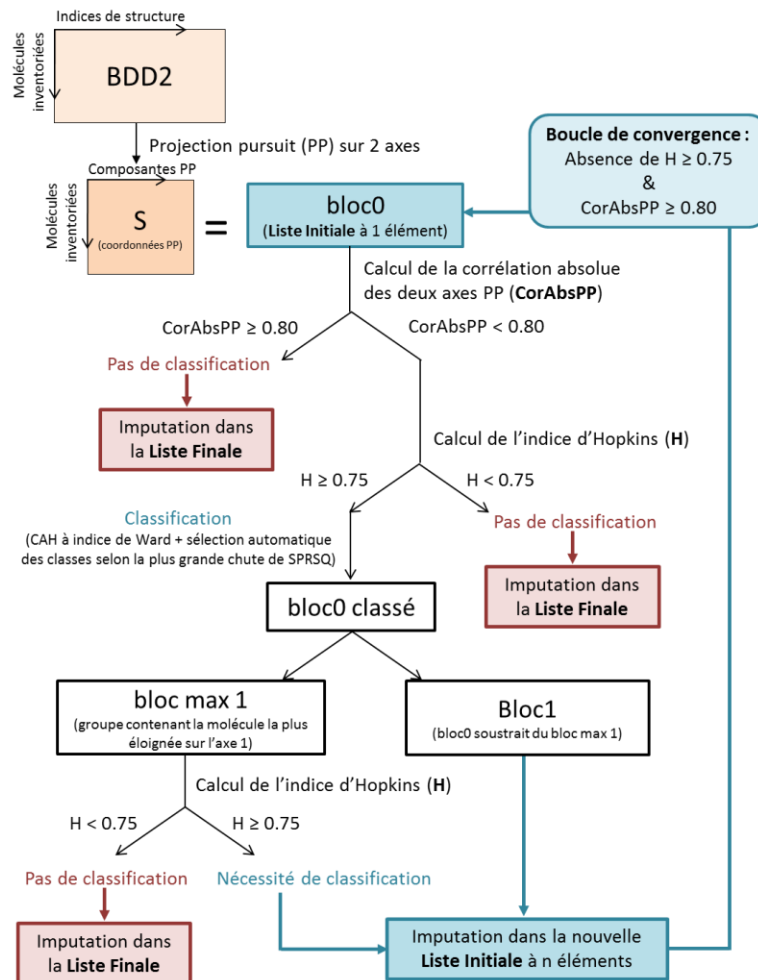


Figure 28: Classification CAH des molécules consolidée par l'indice de Hopkins

La construction des axes par la méthode de « Projection Pursuit » implique que plus une molécule sera loin de l'origine selon l'axe 1, plus celle-ci sera différente des autres. C'est pourquoi les groupes sont constitués de manière itérative afin d'isoler en priorité et au fur et à mesure les groupes de molécules les plus éloignés de l'origine selon l'axe 1, c'est-à-dire les molécules les plus différentes des autres. Ce processus d'itération converge jusqu'à stabilisation de l'axe 1, c'est-à-dire jusqu'à un niveau de corrélation élevé de l'axe 1 avec l'axe 2 (au moins égal à 0,80).

À noter que la classification a été construite sur le jeu de données disponibles et que malgré une automatisation et une généralisation la plus poussée possible, il est éventuel qu'elle soit instable pour d'autres jeux de données (réutilisation de la fonction R dans d'autres cas ou augmentation du nombre de données dans la base de données actuelle de seuils de perception). Par exemple, il se peut que l'automatisation du nombre de classes mène soit à une classification qui ne converge pas assez tôt impliquant une division du jeu de données parfois trop importante allant jusqu'à obtenir des classes contenant une seule molécule.

La classification obtenue permet de choisir le jeu de construction (« train ») qui représente la plus large couverture de structures différentes. Pour cela, le barycentre de chaque classe est calculé dans le but de sélectionner les points les plus éloignés de celui-ci. Par classe, les coordonnées de ce point sont notées « x_{Bary} » pour l'axe 1 et « y_{Bary} » pour l'axe 2 et sont définies selon l'Équation 6.

$$x_{Bary} = \frac{\sum_i^n x_i}{n} \quad y_{Bary} = \frac{\sum_i^n y_i}{n} \quad \begin{array}{l} n = \text{nombre de molécules dans la classe étudiée} \\ i = \text{élément de la classe étudiée } (1 \leq i \leq n) \end{array}$$

Équation 6 : Coordonnées du barycentre d'une classe contenant n molécules

Puis, toujours dans la même optique, par classe, la distance de chaque point (I) au barycentre (B) est calculée (Équation 7).

$$BI = \sqrt{(x_i - x_{Bary})^2 + (y_i - y_{Bary})^2}$$

x_i = Abscisse du point I (sur Axe 1)
 y_i = Ordonnée du point I (sur Axe 1)
 x_{Bary} = Abscisse du point B (sur Axe 2)
 y_{Bary} = Ordonnée du point B (sur Axe 2)

Équation 7 : Calcul de la distance d'un point I ($x_i; y_i$) à un point B ($x_{Bary}; y_{Bary}$)

Le jeu de données de construction du modèle (« train ») est construit de la manière suivante :

1. Sélection des 4 molécules ayant les coordonnées extrémales sur les axes 1 et 2 et des molécules dont le seuil de perception a été mesuré par Olentica ;
2. Tri des molécules restantes selon les ordres de grandeur croissants des seuils de perception en premier niveau et selon leur distance au barycentre de leur classe croissant en deuxième niveau ;
3. Sélection de 3 molécules sur 4 en partant de la deuxième molécule du jeu de données trié (dans le cas des petits groupes où les extremums représentent la moitié du groupe alors seuls ces extremums et les molécules dont le seuil de perception est mesuré par Olentica constituent le jeu de construction).

Le modèle est ensuite calculé à partir du jeu de données de construction ainsi constitué.

3.4.2 Définition des jeux de construction et jeu de validation

Les paragraphes suivants présentent les résultats de la démarche de division du jeu de données en jeu de construction et de validation qui serviront de base au modèle de prédiction des seuils olfactifs.

3.4.2.1 Tri des indices par la méthode du Lasso

Un tri est appliqué par typologie d'indices (entiers, booléens, réels) pour ne retenir que les indices pertinents à la construction du modèle. Le résultat de ce tri est présenté en Tableau 26.

Tableau 26 : Tri des indices par la technique du Lasso

Variables Entières		Variables Booléennes		Variables Réelles	
Toutes (36)	Conservées (26)	Toutes (8)	Conservées (3)	Toutes (24)	Conservées (18)
C	X	P		XLogP3	X
H		Acides	X	MM_Molecule	X
Br	X	Alcanes	X	AMW	
Cl	X	Amides		Ni	X
F		Anhydrides		W.Dh.	X
I		Phosphores		X1Xp	
N	X	Terpènes	X	X0X	X
O	X	Peroxides		X1X	X
S	X			X0Xv	X
NbLtot				X1Xv	X
NbLS				J.Dh.	X
NbLD	X			Jt.Dh.	X
NbLT	X			MTI	X
NbLConj	X			HararyDh	
nCprimaire	X			GUTMAN	X
nCsecondaire				W..Ah.	
nCtertiaire				J..Ah.	X
nCquaternaire	X			Jt..Ah.	X
CYCLE	X			W...CEP.	X
M1	X			J...CEP.	X
M2	X			Jt...CEP.	X
Alcenes	X			SiTot	X
Alcools				ZEP	
Alcynes	X			Electronegativity	
Aldehydes	X				
Amines	X				
Aromatiques	X				
Bromes	X				
Cetones	X				
Chlores	X				
Esters	X				
Ethers.	X				
Fluores					
Iodes					
Soufres	X				
Sulfinyles	X				

Des indices sont éliminés pour les trois types. Le nombre d'indices diminue de soixante-huit à quarante-sept. Les indices restants suffisent pour définir une espace de modélisation qui couvre un large spectre de structure.

3.4.2.2 Analyse de la structuration des données dans l'espace défini par les axes de « Projection Pursuit »

Pour visualiser l'entropie des 350 molécules, celles-ci sont projetées dans l'espace défini par les deux axes principaux de la méthode « Projection Pursuit » en Figure 29. Les coordonnées numériques sont présentées en Annexe 14.

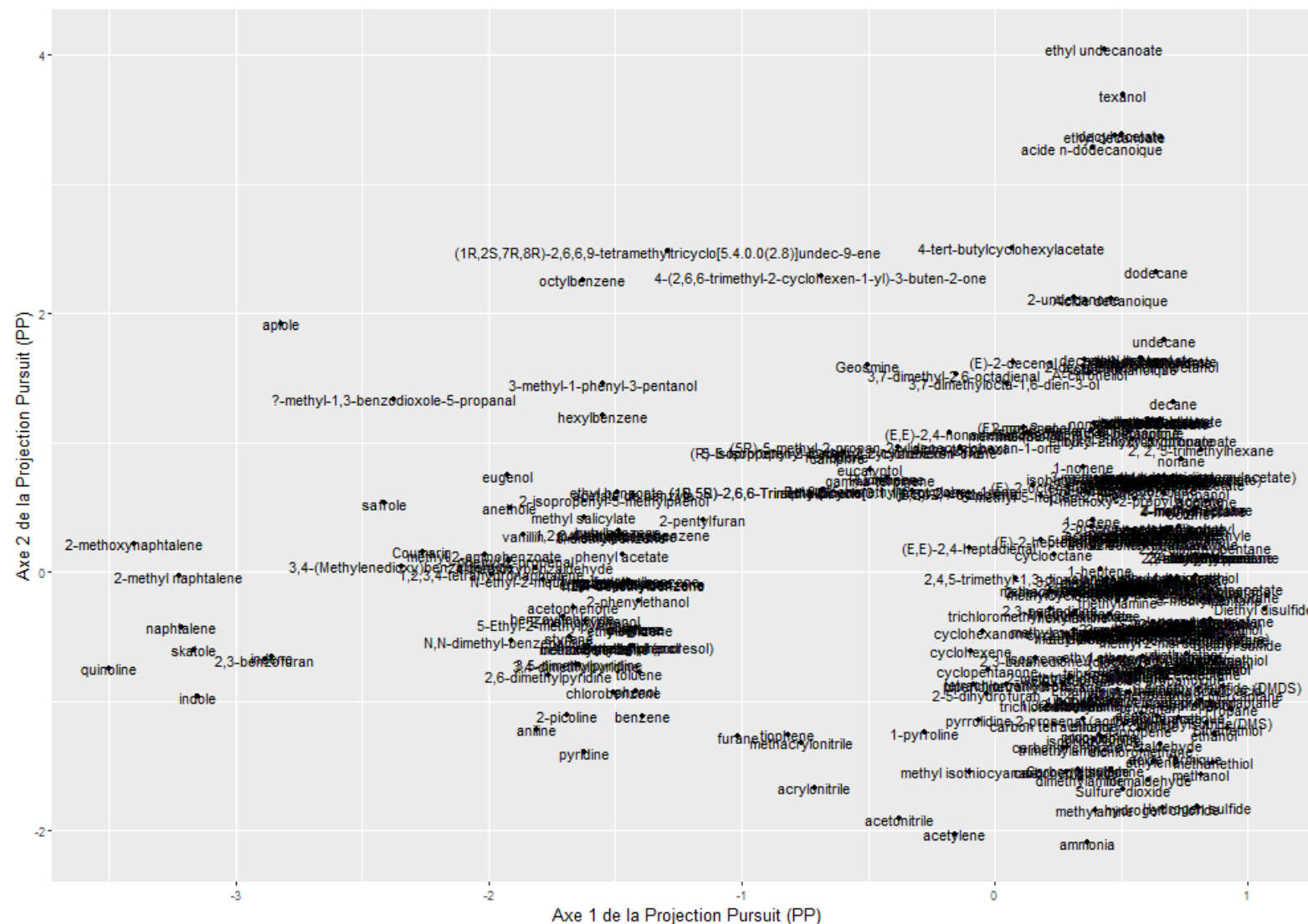


Figure 29: Projection des molécules dans l'espace défini par les axes principaux de la méthode de « Projection Pursuit »

Ce graphe illustre la structuration principale du jeu de données dans l'espace défini par les deux premiers axes de la méthode de « Projection Pursuit ». Il permet de constater une division du jeu de données en au moins deux classes car cette méthode priorise les groupes les plus différents du reste du jeu de données sur l'axe horizontal (Axe 1) (ROBERTS et EVERSON 2001).

La construction des projections nécessite la construction d'une matrice intermédiaire « A » appelée « mixing matrix ». Cette matrice contient les coordonnées des indices sur les axes de la « Projection Pursuit » (soit les « loadings » des indices) [(ADALI et al. 2009) ; (HYVARINEN 2010)]. Cela permet de projeter également les indices sur les axes de la projection (Figure 30). Les coordonnées numériques du positionnement de ces variables sont présentées en Annexe 15.

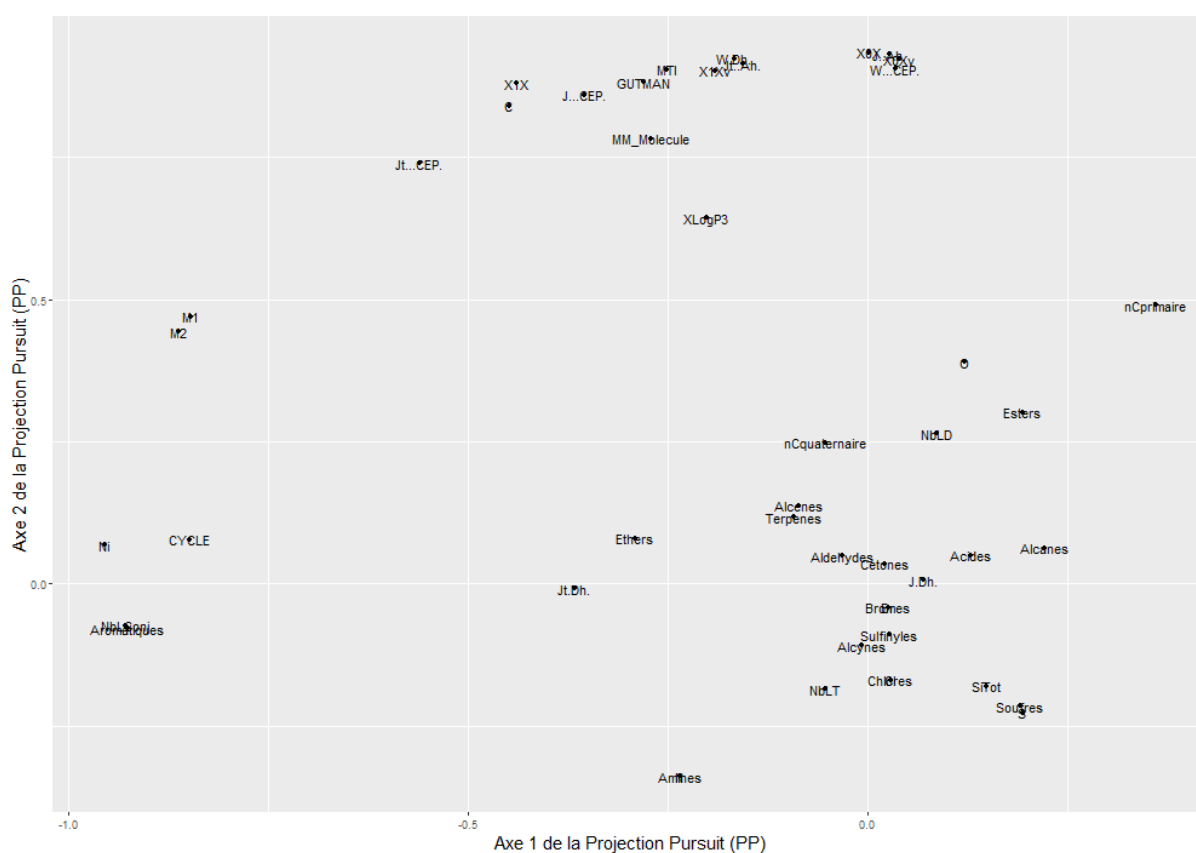


Figure 30: Projection des indices dans l'espace défini par les axes principaux de la méthode de Projection Pursuit

La dispersion des indices est moins étendue que celle des molécules sur les axes principaux de la « Projection Pursuit ». La Figure 30 permet de compléter la Figure 29 par l'identification des indices expliquant le plus la séparation des molécules. Par exemple, les indices « CYCLE », « Ni », « Aromatiques » et « NbLConj » (définis en Tableau 25) apparaissent comme les indices isolant toutes les molécules à gauche de l'axe 1. Cette information est cohérente avec la nature des molécules

concernées à gauche en Figure 29. En résumé, il semblerait que l'axe 1 discrimine les molécules selon des indices entiers/booléens tels que la nature des groupements fonctionnels présents ou le nombre de cycles en opposition à l'axe 2 qui différencie les molécules selon des indices réels tels que la masse molaire.

La Figure 31 présente, sur ces mêmes axes, les molécules, catégorisées par ordre de grandeur de leur seuil de perception. Les ordres de grandeurs sont définis comme étant la partie entière du logarithme décimal des valeurs de seuils de perception.

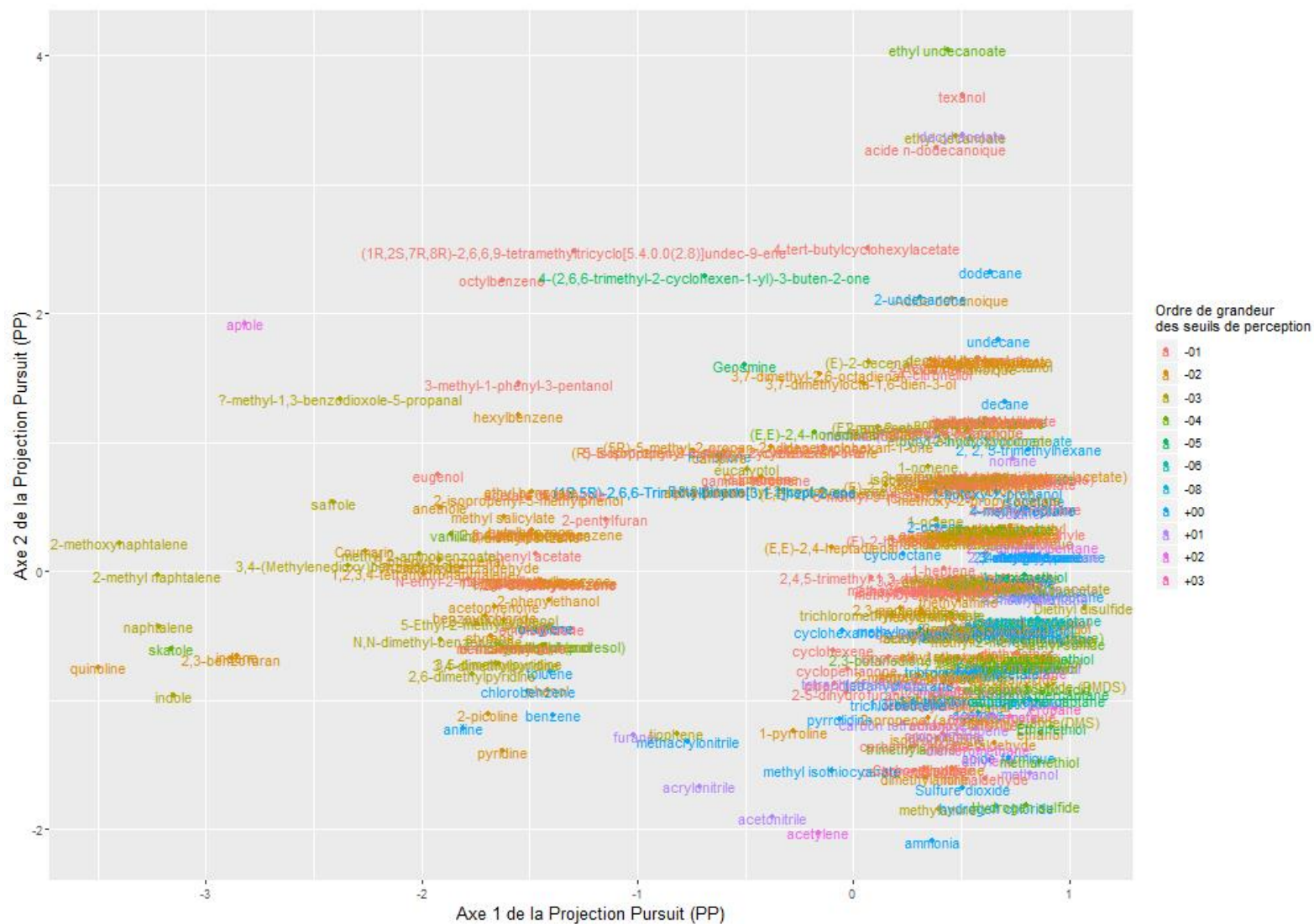


Figure 31 : Projection des molécules dans l'espace défini par les axes principaux de la Projection Pursuit

Cette figure montre que les molécules ne sont pas regroupées en fonction de la valeur de leur seuil de perception. Aucune corrélation ne peut être directement mise en évidence entre la position des molécules sur les axes de « Projection Pursuit » et l'ordre de grandeur des seuils de perception. Une augmentation du nombre de projections (quatre) n'apporte pas une amélioration de cette corrélation.

3.4.2.3 Classification des molécules et description des classes

L'application de la classification, expliquée en paragraphe 3.4.1, permet de constituer huit classes détaillées en Annexe 16. Avant de décrire ces classes, la stabilité de la classification est vérifiée via le calcul de l'indice de Hopkins de la fusion de chacune des classes deux par deux (Tableau 27).

Tableau 27 : Indice de Hopkins de la fusion des classes deux par deux

	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8
Classe 1	0,00	0,83	0,71	0,63	0,88	0,85	0,81	0,91
Classe 2	0,83	0,00	0,88	0,95	0,77	0,80	0,64	0,92
Classe 3	0,71	0,88	0,00	0,68	0,91	0,81	0,66	0,85
Classe 4	0,63	0,95	0,68	0,00	0,92	0,77	0,92	0,90
Classe 5	0,88	0,77	0,91	0,92	0,00	0,84	0,76	0,92
Classe 6	0,85	0,80	0,81	0,77	0,84	0,00	0,79	0,82
Classe 7	0,81	0,64	0,66	0,92	0,76	0,79	0,00	0,86
Classe 8	0,91	0,92	0,85	0,90	0,92	0,82	0,86	0,00

En orange sont surlignés les indices de Hopkins inférieurs à la limite de classabilité 0,75. La valeur la plus éloignée de 0,75 est 0,63. Cette valeur reste néanmoins très proche de 0,75 et traduit une stabilité de la classification.

La décomposition de ces classes par familles chimiques est présentée en Figure 32. Cette décomposition est également établie par groupements fonctionnels (Annexe 17).

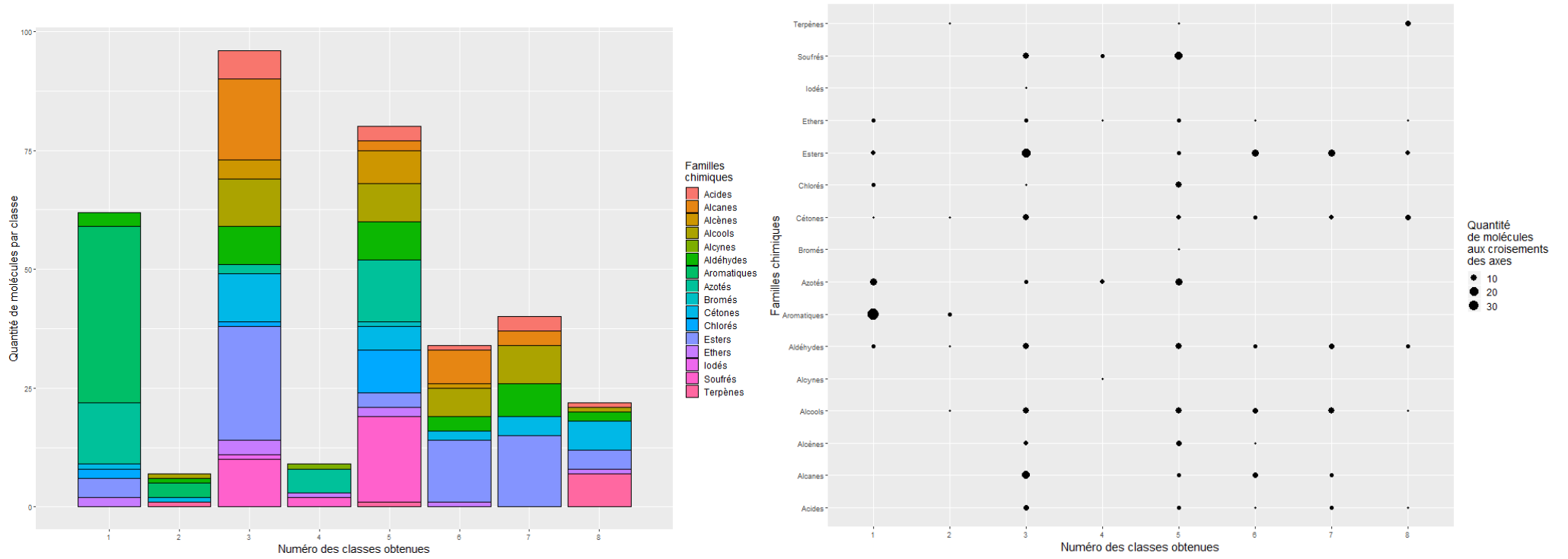


Figure 32 : (a) Histogramme de la composition des classes en familles chimiques (b) et répartition des familles chimiques dans ces classes

Les Figure 32a et Figure 32b montrent que la taille des classes est hétérogène et qu'il n'y a pas de catégorisation par famille chimique. Par exemple, les classes contenant peu de molécules, telles les classes 2 ou 4, ne correspondent pas à une seule famille chimique. Cependant, la classe 1 est composée majoritairement de molécules aromatiques et la plupart des molécules aromatiques présentes dans la base (93%) sont par ailleurs réunies dans cette classe 1. Cette observation est également vérifiée pour la famille des terpènes dans la classe 8 (78%). Or, comme les classes sont réalisées selon un ordre de distinction des molécules les unes par rapport aux autres, cela illustre clairement que la famille majoritairement isolée en premier est celle des aromatiques (classe 1) et que la famille majoritairement résiduelle est celle des terpènes (classe 8).

D'autres familles chimiques se trouvent également majoritairement dans une seule classe comme cela est le cas des molécules iodées (classe 3), des composés bromés (classe 5), ou des alcynes (classe 4). Les molécules azotées sont présentes exclusivement dans les classes 1, 3, 4 et 5 mais les autres familles chimiques comme les esters, les aldéhydes et les cétones sont réparties de manière relativement homogène dans toutes les classes. La taille des familles de composés iodés, bromés et des alcynes dans le jeu de données peut expliquer leur appartenance à une seule classe.

Les molécules sont maintenant projetées dans l'espace défini par les axes principaux de la méthode « Projection Pursuit » par classes (Figure 33).

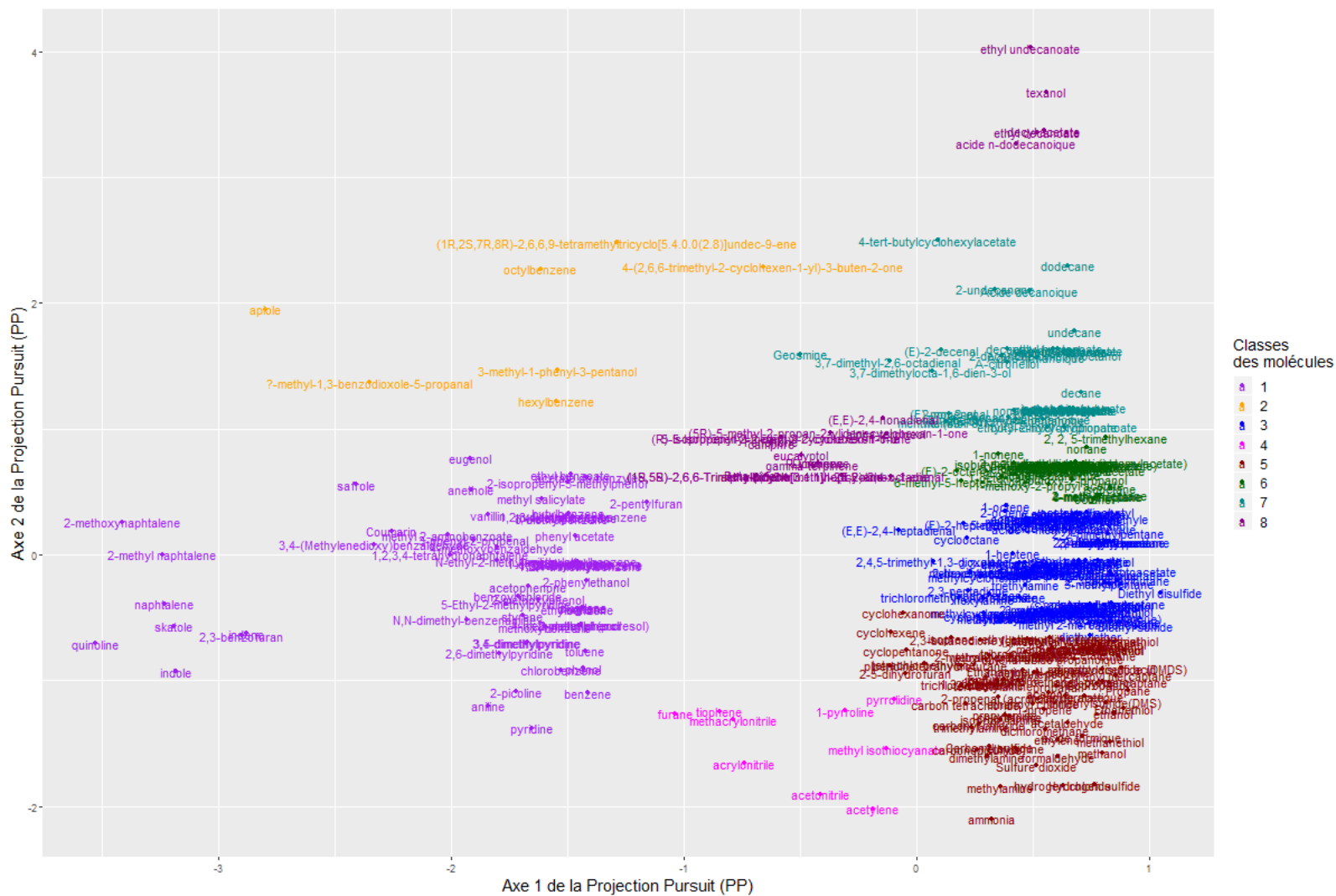


Figure 33 : Projection des molécules par classe sur l'espace défini par les axes principaux de la « Projection Pursuit »

En première analyse, la Figure 33 confirme l'existence de deux classes décrites selon l'axe 1. Il est à noter que les molécules les plus extrêmes sur l'axe 1 (à la gauche du graphique) sont regroupées avec le groupe central. Ce groupe de molécule n'est donc pas divisé au sens de la classification appliquée.

Dans un deuxième temps, la classification permet de mettre en évidence la division des molécules sur l'axe 2. Ce dernier aspect est plus difficile à interpréter, mais, confronté aux résultats illustrés par la Figure 30, il permet de distinguer les molécules azotées (répartis sur les classes 1, 4 et 5) en bas de l'axe 2 de celles ayant un encombrement stérique important en haut de l'axe 2 (réparties sur les classes 2, 7 et 8). Cette figure permet de confirmer une meilleure séparation des molécules selon l'axe 1 que selon l'axe 2.

Afin de construire notre modèle, chaque classe définie dans ce paragraphe est divisée en un jeu de données de construction et un jeu de données de validation comme présenté ci-dessous.

3.4.2.4 Définition du jeu de construction

La division du jeu de données en un jeu de données de construction (« train ») et un jeu de données de validation (« test ») comme définis dans le paragraphe 3.4.1 est illustrée pour une classe contenant un grand nombre de molécules (classe 3) en Figure 34 et une classe contenant un petit nombre de molécules (classe 2) en Figure 35.

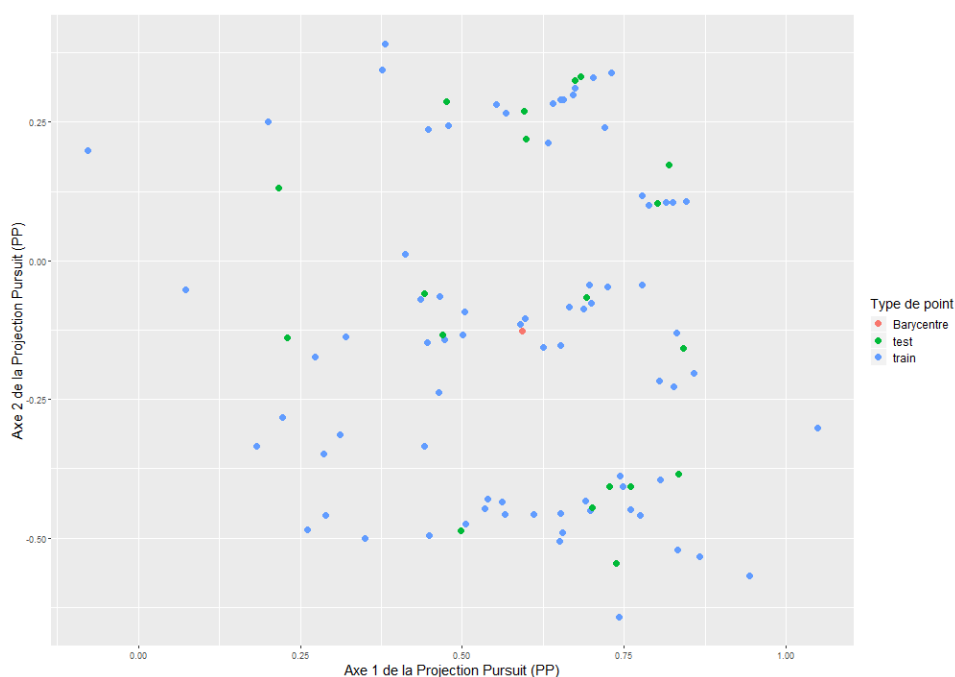


Figure 34 : Division du jeu de données en données de construction (« train ») et en données de validation (« test ») pour la classe 3 (96 molécules)

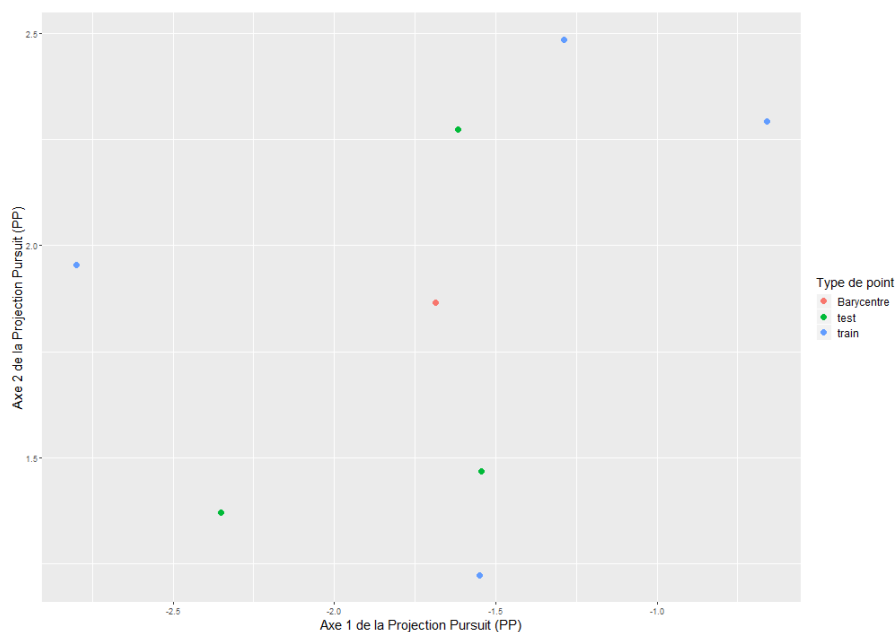


Figure 35 : Division du jeu de données en données de construction (« train ») et en données de validation (« test ») pour la classe 2 (8 molécules)

Ces deux figures permettent d’observer que la sélection du jeu de données de construction couvre l’espace défini pour chaque classe.

La répartition de la division du jeu de données global en jeu de données de construction et en jeu de données de validation sur les axes de la méthode de « Projection Pursuit » est présentée en Figure 36.

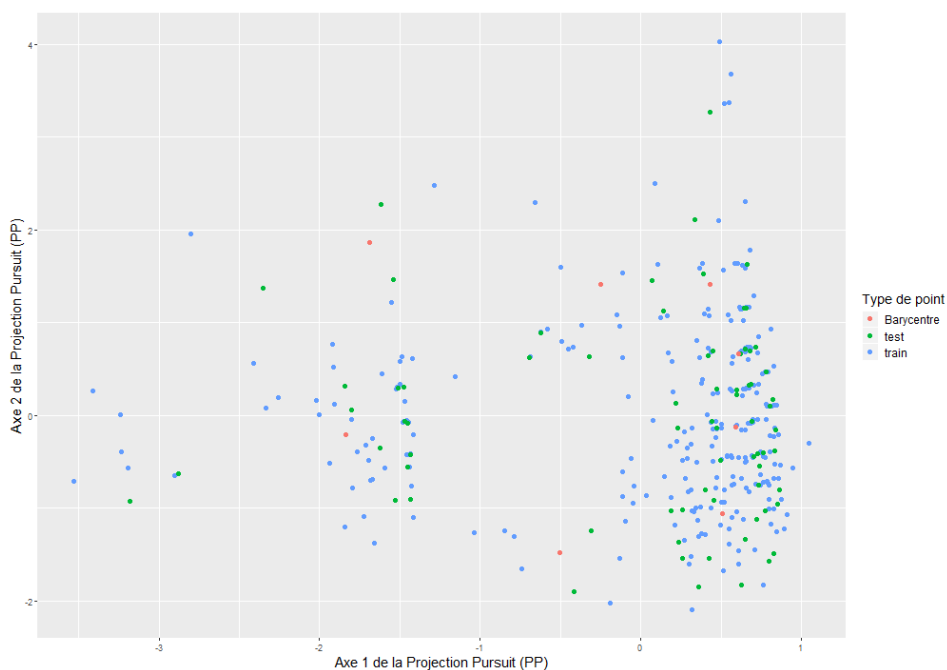


Figure 36 : Division des 350 molécules en données de construction (« train ») et en données de validation (« test »)

Cette figure montre que l'approche adoptée permet d'obtenir un bon recouvrement de l'espace défini par les propriétés des molécules. La démarche amène à diviser le jeu de données en 277 molécules de construction et 73 molécules de validation. Comme expliqué également au paragraphe 3.4.1, cette division couvre également la diversité des ordres de grandeur de seuils de perception des molécules de la base de données.

Ces jeux de données de construction et de validation ainsi constitué vont nous permettre de générer un modèle de prédiction des seuils de perception.

3.5 Génération du modèle

Une fois l'approche de modélisation déterminée, les variables d'entrées calculées et triées et la division du jeu de données établie, la technique de modélisation peut alors être sélectionnée. Celle-ci tiendra compte de la quantité des données, de leur format et de leur nature.

3.5.1 Méthodologie de modélisation

La méthodologie de modélisation est divisée en trois parties consistant en un prétraitement des données à modéliser, suivi de la modélisation à proprement parlée et enfin par une évaluation du modèle.

3.5.1.1 Prétraitement de la variable réponse de seuil de perception

Comme cela a pu être mentionné au chapitre 2, la répartition des ordres de grandeur des valeurs de seuils de perception à modéliser est très hétérogène. La Figure 37 illustre cette hétérogénéité.

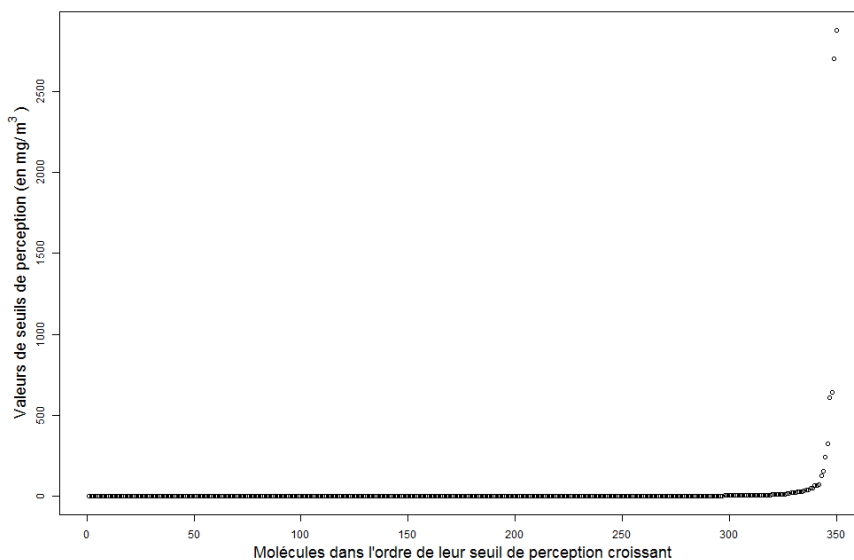


Figure 37 : Valeurs des seuils de perception à modéliser dans l'ordre croissant

Cette figure confirme la grande étendue de la gamme des seuils de perception et plus particulièrement l'hétérogénéité de la représentativité des valeurs. La distribution particulière de cette variable la rend

difficile à modéliser. Il est donc choisi de linéariser cette variable par transformation en logarithme décimal (\log_{10}) présentée en Figure 38.

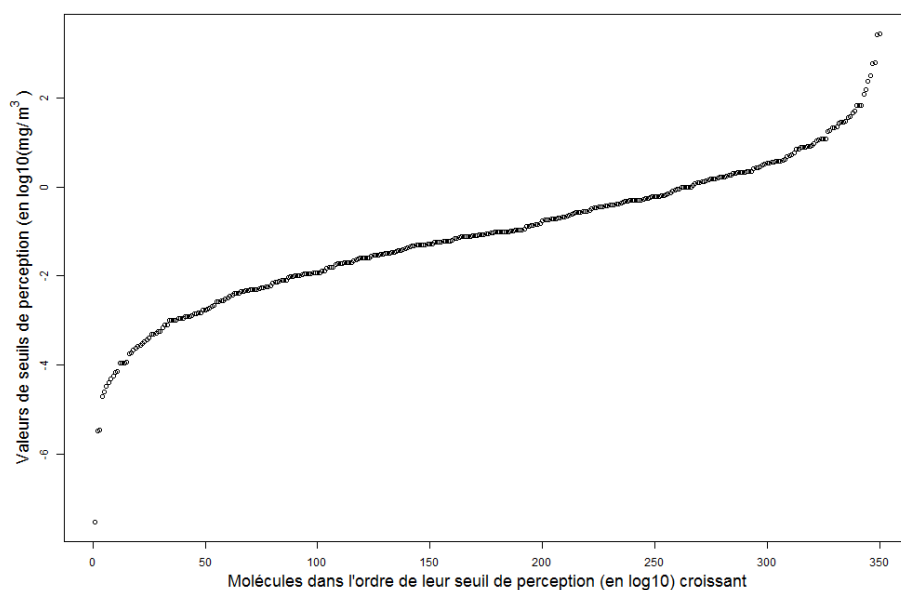


Figure 38 : Valeurs de seuils de perception (en \log_{10}) à modéliser classées dans l'ordre croissant

Cette transformation permet alors une modélisation moins précise mais plus robuste car plus aisée à modéliser en raison de l'homogénéité des valeurs. Cela permet de proposer une approche limitant le sur-apprentissage et aussi d'augmenter la robustesse du modèle au vue de la qualité des données d'entrée.

3.5.1.2 Modèle global : Choix et application du traitement statistique

Le but de cette partie est de définir le modèle le plus adapté aux données à mettre en relation. Suite à la division du jeu de données, les blocs de données à mettre en relation sont présentés en Figure 39.

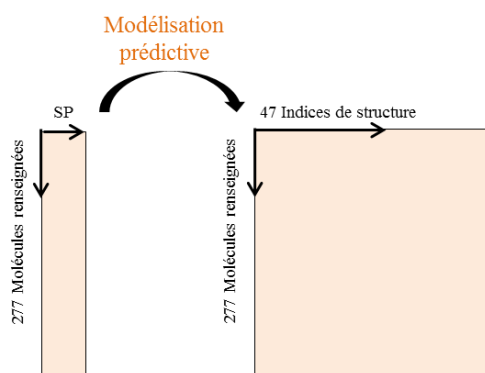


Figure 39 : Données à mettre en relation pour le modèle global

Il faut donc construire un modèle à partir de 277 molécules et 47 indices de structure et le valider sur 73 molécules. L'un des points positifs est que le nombre de molécules n'est pas un point limitant puisqu'il est supérieur au nombre de variables.

La répartition des molécules sur l'axe 1 défini par la Projection Pursuit, présentée au paragraphe 3.4.2.2, encourage à appliquer un modèle non linéaire [(DUROCHER 2015) ; (DUROCHER, CHEBANA, et OUARDA 2015)]. Par conséquent, un traitement statistique permettant de construire un modèle non linéaire tout en évitant le plus possible le sur-apprentissage est privilégié.

L'approche choisie pour cette modélisation est la régression « Support Vecteur Machine » ou SVM. Il s'agit d'une technique mise en place en 1992 par Boser (BOSER, GUYON, et VAPNIK 1992) puis corrigée et de nouveau publiée par Vapnik en 1995 (VAPNIK 1995). Il a été choisi d'appliquer cette approche car elle combine la notion de marge maximale et de fonction noyau en faisant appel à l'apprentissage automatique (CORNUEJOLS 2002). Par conséquent, elle permet d'appliquer des modèles non-linéaires tout en optimisant la précision des prévisions en évitant le sur-apprentissage (IBM 2012).

L'un des autres points positifs de la SVM est qu'elle permet d'identifier les molécules « supports » à la modélisation. Ces molécules « supports » correspondent aux molécules permettant de tracer l'hyperplan de modélisation. Celui-ci est construit de manière à maximiser sa distance minimale aux échantillons d'apprentissage (CORNUEJOLS 2002). La fiabilité du seuil de perception de ces molécules « supports » impactera davantage la fiabilité du modèle que le seuil de perception des autres molécules (CORNUEJOLS 2002). La mesure du seuil de perception de ces molécules par Olentica pourra alors être priorisée par rapport aux autres molécules de la base afin d'obtenir une plus grande fiabilité du modèle

D'un point de vue pratique, la SVM est une technique aisée et rapide d'utilisation qui nécessite moins de paramètres que d'autres techniques non linéaires comme par exemple les réseaux de neurones (CORNUEJOLS 2002). Elle tend à remplacer ces derniers dans différents domaines d'application comme l'ingénierie, la recherche d'information ou encore la bio-informatique (SCHOLKOPF et SMOLA 2002).

Dans ce projet, l'approche SVM est appliquée à l'aide de la fonction « ksvm » du package « kernlab » (SCHOLKOPF et SMOLA 2002). L'utilisation d'un noyau gaussien a été choisie dans notre cas. L'un des paramètres déterminants du modèle est la fonction de perte insensible appelée « epsilon » mesurant la distance entre la réponse observée et la réponse prédite (CORNUEJOLS 2002). Par défaut dans la fonction « ksvm », sa valeur est fixée à 0,1 mais cette valeur n'est pas toujours celle qui permet l'obtention du meilleur modèle. Pour choisir ce paramètre de manière optimale il faut que l'erreur de prédiction sur le jeu de construction (« train ») et sur le jeu de validation (« test ») soient similaires afin

d'éviter un sur-apprentissage. Pour cela, le modèle est lancé pour une valeur d'épsilon variant de 0,1 à la plus grande valeur d'épsilon permise par le jeu de données avec un pas de 0,1.

L'erreur mesurée du modèle correspond à l'écart-type des erreurs individuelles (Équation 8).

$$\text{Erreur} = \sqrt{\frac{\sum_{i=1}^N (x_{pred\ i} - x_{mes\ i})^2}{N-1}} \quad \text{où}$$

N = nombre de valeurs prédites
 x_{pred} = la valeur prédite
 x_{mes} = la valeur initiale

Équation 8 : Calcul de l'erreur d'un modèle

La valeur minimale d'épsilon permettant une erreur similaire sur la modélisation du jeu de construction (« train ») et celle du jeu de validation (« test ») est donc choisie comme valeur pour l'établissement de notre modèle.

3.5.1.3 Évaluation du modèle global par validation croisées

Afin de vérifier que le modèle établi ne présente pas de sur-apprentissage, l'erreur de prédiction est calculée par validation croisée selon la technique « k-fold » aléatoire [(OBOZINSKI 2013) ; (AZENCOTT 2017) ; (ARLOT 2017)]. Cela permet de calculer l'erreur de prédiction globale du modèle pour différentes divisions du jeu de données aléatoires mais assurant de construire et de valider le jeu de données au moins une fois sur toutes les valeurs. Le k est ici choisi à 10 car c'est une pratique couramment retrouvée dans la littérature [(AZENCOTT 2017) ; (ARLOT 2017)] et que cette valeur est un bon compromis entre « biais » et « variance » (RAKOTOMALALA 2008). Cette valeur permet également de diviser les 350 molécules de manière uniforme, ce qui est un critère important de la division (OBOZINSKI 2013).

3.5.2 Résultats de la modélisation

Cette partie décrit la SVM construite et les résultats associés.

3.5.2.1 SVM et modèle global

La première étape de l'application de SVM est le choix de la fonction de perte « epsilon ». L'évolution de l'erreur sur le jeu de données de construction et le jeu de données de validation en fonction de la valeur d'épsilon est présentée en Figure 40.

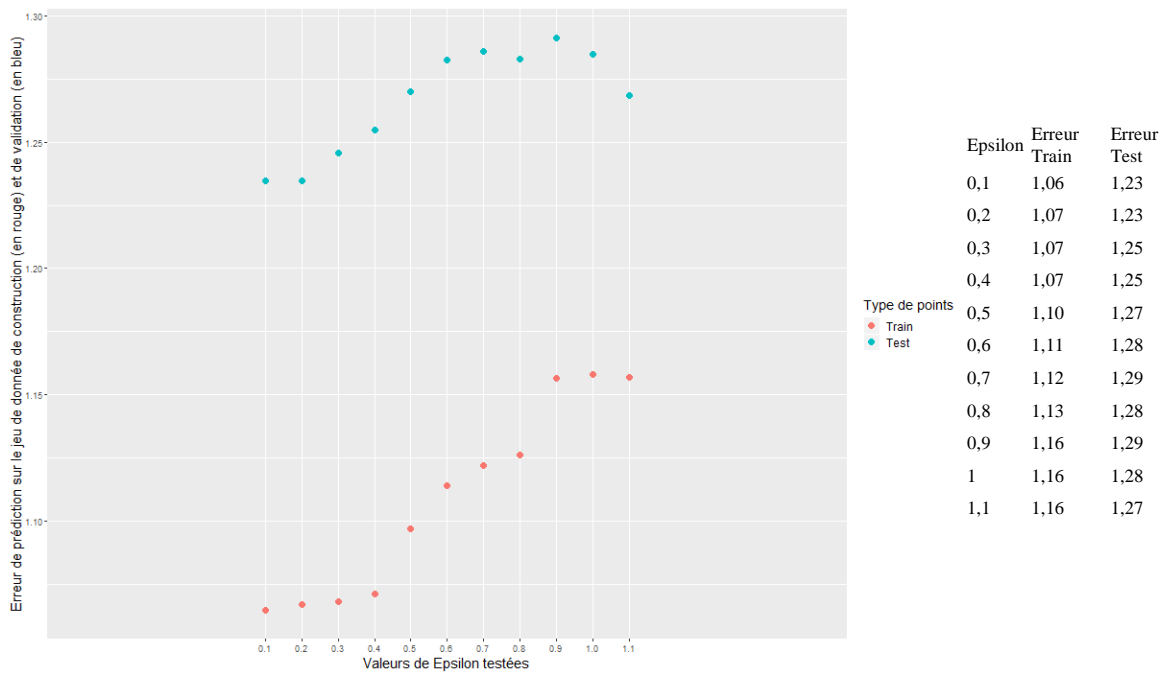


Figure 40 : Erreurs calculées sur le jeu de construction (« train ») et sur le jeu de validation (« test ») en fonction de la valeur d’epsilon

Choisir une valeur d’epsilon permettant des erreurs sur le jeu de construction et sur le jeu de validation similaires et minimales amène à choisir la valeur de 0,1. Cette valeur entraîne une erreur sur le jeu de construction de 1,06 et sur le jeu de validation de 1,23. De plus, la similarité de ces valeurs permet d’éviter un sur-apprentissage du modèle.

La valeur d’epsilon choisie permet d’appliquer la fonction SVM de manière optimale. Le graphique des valeurs prédites en fonction des valeurs observées est présenté en Figure 41. Les données numériques sont présentées en Annexe 18.

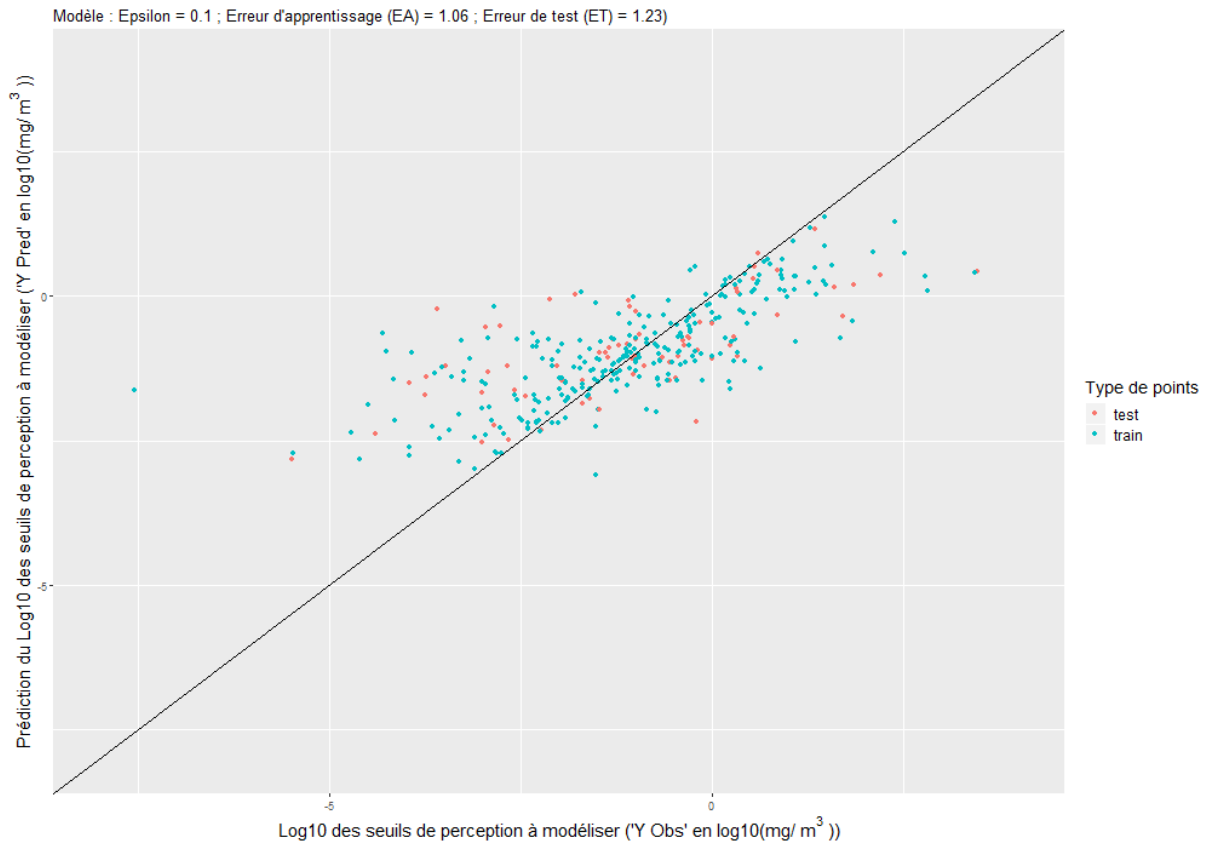


Figure 41 : Graphe des valeurs prédites en fonction des valeurs observées (350 molécules)

Le coefficient de corrélation est de 0,76 pour le jeu de données de construction et de 0,70 pour le jeu de données de validation. Ces valeurs de coefficient de corrélation sont encourageantes car elles valident un modèle qui produit des données avec une erreur raisonnable. De plus la Figure 41 souligne un apprentissage équilibré et non sur appris en raison de la répartition homogène des données de construction et des données de validation autour de l'axe central qui représente une prédiction exacte des points.

L'examen de la Figure 41 permet également de constater que l'erreur maximale du modèle semble être concentrée sur une zone de valeur de logarithme décimal des seuils de perception comprise entre -5 et -2,5 malgré des valeurs de seuils de perception correctement prédites par le modèle dans cette zone. Dans cette zone, la prédiction surévalue le seuil, et cet écart de prédiction concerne aussi bien des données de construction que des données de validation. De plus, cette zone correspondant aux seuils de perception les plus bas et par conséquent aux molécules à l'origine des impacts les plus importants sur l'odeur, il conviendra de porter une attention particulière aux valeurs obtenues.

Une autre zone plus étendue, également observable en Figure 41, présente une sous-estimation de la valeur du seuil de perception. Elle concerne des molécules dont le logarithme décimal des seuils de perception est compris entre -1 et 3. Cette zone est moins critique pour le modèle car elle concerne des

molécules dont le seuil de perception est plus élevé et qui ont par conséquent moins d'impact sur la concentration de l'odeur.

Nous avons donc choisi de cibler les molécules présentant de forts écarts de prédiction pour tenter d'identifier des structures communes pouvant expliquer cet écart.

Avant de poursuivre, il est nécessaire de vérifier l'erreur de prédiction globale du modèle par validation croisée.

3.5.2.2 Résultat de la validation croisée

Les résultats de la validation croisée sont présentés en Tableau 28.

Tableau 28 : Résultat des erreurs par validation croisée

Sous-échantillons	Erreur Train	Erreur Test
Sous-échantillonnage 1	1,04	1,35
Sous-échantillonnage 2	1,07	1,12
Sous-échantillonnage 3	1,08	1,03
Sous-échantillonnage 4	1,08	1,05
Sous-échantillonnage 5	1,08	0,99
Sous-échantillonnage 6	1,10	1,15
Sous-échantillonnage 7	1,09	1,16
Sous-échantillonnage 8	1,05	1,31
Sous-échantillonnage 9	1,06	1,21
Sous-échantillonnage 10	1,04	1,80
Moyenne	1,07	1,22

Le Tableau 28 montre qu'en moyenne, l'erreur sur les données de construction est de 1,07 et celle sur les données de validation de 1,22. Ces valeurs sont similaires à celles obtenues avec le modèle global. Cela permet de montrer que la division stable du jeu de données proposée a été correctement appliquée.

3.5.2.3 Les molécules supports

L'approche SVM identifie 236 molécules « supports » sur les 277 molécules qui ont servi à construire le modèle. Ces molécules sont présentées en Annexe 19. Ce nombre important de molécules s'explique par la répartition des molécules dans l'ensemble de l'espace défini par la méthode « Projection Pursuit ». Améliorer la fiabilité du modèle passerait par la mesure du seuil de perception de ces 236 molécules en priorité. Étant donné le caractère très chronophage et couteux du protocole de mesure, cette quantité de molécules est encore trop importante. L'un des points positifs est que 39 des 45 molécules dont le seuil de perception a été mesuré par Olentica sont comprises dans les 236 molécules « supports ». Cela contribue à la fiabilité du modèle.

3.6 Affinage du modèle global par analyse des erreurs individuelles

Comme souligné précédemment (§3.5.2.1), une attention particulière doit être portée aux molécules dont l'erreur de prédiction du seuil de perception est maximale pour déterminer si une structure particulière permet d'expliquer ces écarts. Cette analyse est d'autant plus importante que le seuil de perception est faible. C'est pourquoi, une méthodologie d'analyse des erreurs de prédiction est proposée.

3.6.1 Méthodologie d'analyse

Ce paragraphe présente la méthodologie permettant d'affiner le modèle global par analyse des erreurs individuelles.

3.6.1.1 Analyse exploratoire des erreurs

Une première partie de l'analyse des erreurs consiste à adopter une démarche exploratoire dans le but d'identifier l'existence éventuelle d'une relation entre la structure des molécules et l'erreur de prédiction de leur seuil de perception par le modèle global.

Pour cela, les erreurs de prédiction sur le logarithme des seuils de perception des 350 molécules sont arrondies à l'unité. Cela permet de comptabiliser le nombre de molécules par ordre de grandeur et de rechercher l'existence d'une cohérence de structure au sein de chaque catégorie d'ordre de grandeur.

Un affinage par modélisation locale est proposé par la suite pour compléter ces observations.

3.6.1.2 Analyse prédictive des erreurs

Cet affinage local consiste à diviser l'ensemble du jeu de données en sous-domaines et à proposer un modèle local pour chacun de ces sous-domaines. Ces sous-domaines sont déterminés à partir des résultats de l'analyse exploratoire précédente des erreurs de prédiction du logarithme décimal des seuils de perception des molécules. Les intervalles d'erreur (ou sous-domaines) définis sont présentés dans le Tableau 29.

Tableau 29 : Intervalles des erreurs (log10) permettant de décomposer le jeu de données de molécules en sous-domaines pour l'affinage local du modèle global.

	Catégorie 1	Catégorie 2	Catégorie 3	Catégorie 4
Intervalle des erreurs de prédiction	[0 ; 0,5 [[0,5 ; 1[[1 ; 1,5 [1,5≤

Pour proposer un modèle local pour chacun de ces sous-domaines, chaque molécule est assimilée à un signal discontinu construit par un ensemble d'indices de structure. Or, l'erreur de prédiction du seuil de perception de certaines molécules montre qu'une partie du signal n'est pas expliquée par le modèle global. C'est pourquoi, une décomposition des signaux en sources indépendantes est appliquée. Cela

permet par la suite d'appliquer un modèle additif prenant en compte toutes les structures sous-jacentes pour reconstituer l'information. Cette démarche se base sur les études menées par Mallat qui explique que toute structure peut être reconstruite à partir de l'invariant, ici les sources indépendantes [(BRUNA et MALLAT 2013) ; (MALLAT 2015)].

Pour cela, l'analyse en composantes indépendantes (ACI) est utilisée. Cette technique repose sur la même fonction de perte que la méthode de « Projection Pursuit » à la différence que la construction des axes n'est pas successive mais parallèle [(HYVARINEN et OJA 2000) ; (MARCHINI et al., 2017b)]. Cette construction permet de décomposer un ensemble de variables en sources indépendantes et s'applique à de nombreux domaines tels que les données spectrales, les données multi dimensionnelles ou encore les données analytiques et sensorielles (RUTLEDGE 2016).

La fonction R utilisée est la même que celle utilisée pour la méthode de « Projection Pursuit » mais avec un paramétrage « parallel » pour l'entrée «alg.typ » (MARCHINI et al., 2017b).

La fonction SVM est ensuite appliquée par sous-domaine d'erreurs pour établir un lien entre les composantes indépendantes obtenues suite à l'analyse en composantes indépendantes et l'erreur de prédiction. Étant donné le caractère linéaire conféré par l'indépendance des composantes, il est maintenant recommandé d'appliquer la fonction SVM avec un noyau polynomial d'ordre 2. Malgré ce caractère linéaire, l'application de la SVM est néanmoins préférée à une régression linéaire classique pour limiter le sur apprentissage.

Étant donnée l'indépendance des composantes obtenues suite à l'analyse en composantes indépendantes, le nombre de composantes à retenir n'est pas régi par une démarche adaptée à toutes natures de données (RUTLEDGE 2016). En effet, dans le cas des spectres continus, le critère de Durbin-Watson (DW) a été développé et appliqué avec succès [(BOUVERESSE JOUAN-RIMBAUD et al. 2012) ; (AMMARI, BOUVERESSE JOUAN-RIMBAUD, et RUTLEDGE 2012)]. Il s'agit du rapport entre le signal reconstitué par les composantes et le bruit des données calculé pour une variation du nombre de composantes (BOUVERESSE JOUAN-RIMBAUD et al. 2012). Cependant, ce critère n'a pas été pertinent sur nos données. C'est pourquoi, le critère choisi pour nos données a été adapté de ce critère de Durbin-Watson : la matrice initiale est reconstituée à partir d'un nombre de composantes croissants et la matrice résiduelle correspondante est calculée. La somme des éléments de chaque matrice résiduelle est calculée et le nombre de composantes retenues correspond à la somme résiduelle minimale.

Cette démarche mène à obtenir un modèle par sous-domaine soit quatre modèles locaux dépendant de l'erreur de prédiction du seuil de perception par le modèle global.

3.6.2 Résultat de l'affinage du modèle global par analyse des erreurs

Cette partie permet de présenter les résultats de l'analyse exploratoire des erreurs de prédiction du modèle global ainsi que du travail de prédiction locale des erreurs par sous-domaine.

3.6.2.1 Analyse exploratoire des erreurs

La projection des molécules dans l'espace défini par les deux axes principaux de la « Projection Pursuit » selon la valeur de l'erreur (arrondie à l'unité près) de la prédiction de leur seuil de perception par le modèle globale, est présentée en Figure 42.

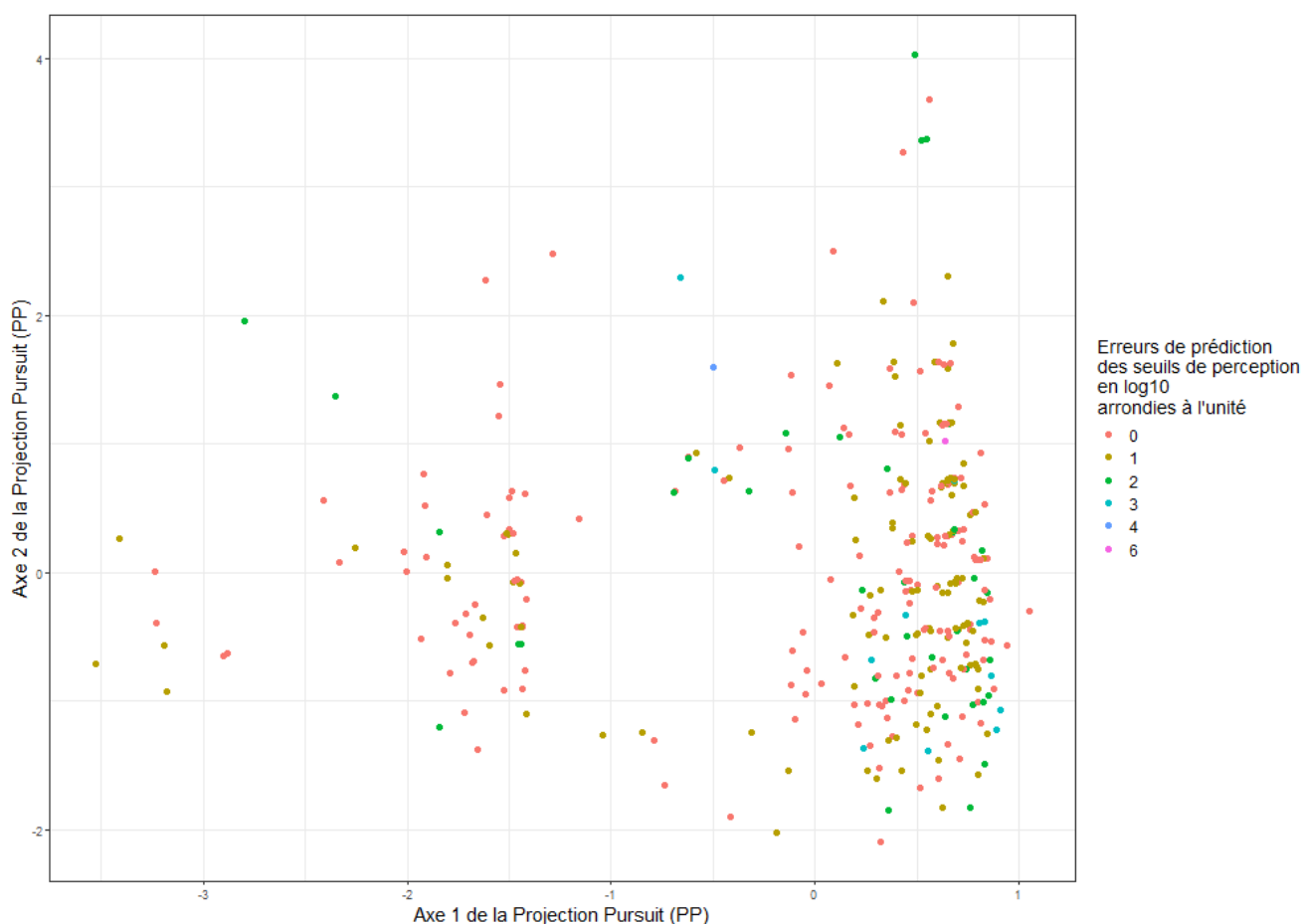


Figure 42 : Projection des molécules sur les axes de « Projection Pursuit » selon l'erreur de prédiction de leur seuil de perception par le modèle global (arrondie à l'unité près)

Ce graphique ne fait pas apparaître de zones où se concentreraient des molécules dont le seuil de perception présenterait une erreur de prédiction importante, mais plutôt une répartition homogène sur les 2 axes de « Projection Pursuit ».

La répartition, par le modèle global, des molécules sur l'axe 1 de « Projection Pursuit » selon l'ordre de grandeur de l'erreur de prédiction de leur seuil de perception (arrondie à l'unité près) est présentée en Figure 43.

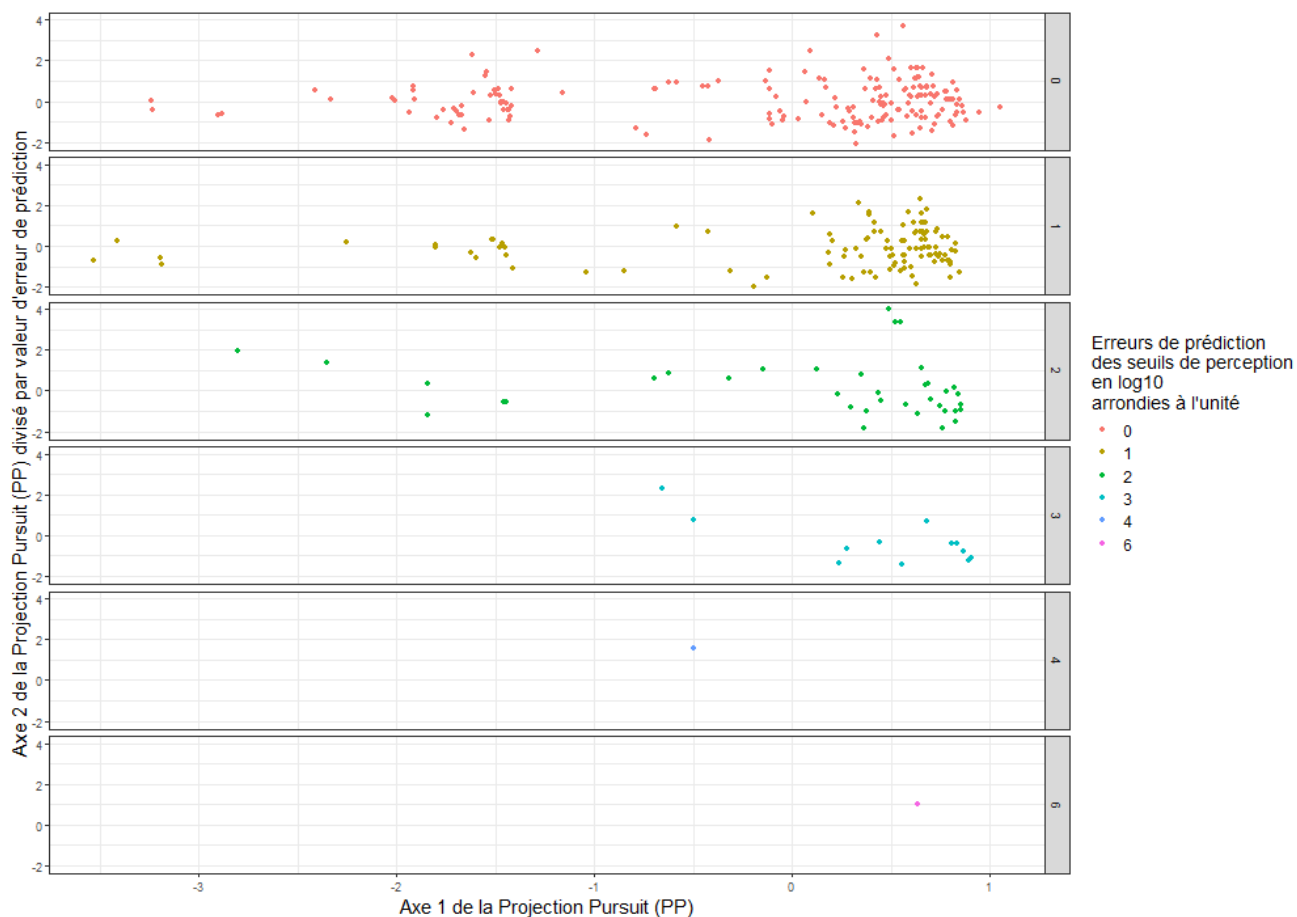


Figure 43 : Répartition des molécules sur l'axe 1 de « Projection Pursuit » selon l'erreur de prédiction par le modèle global

Cette présentation permet d'observer que les molécules dont le seuil de perception est le moins bien prédit sont principalement situées dans la zone la plus dense de l'axe 1. Le manque d'information de construction du modèle ne peut par conséquent pas être incriminé comme responsable de cette mauvaise prédiction.

Un point positif est que l'erreur est inférieure à 1 pour 254 molécules, et qu'elle est inférieure à 1,5 pour 297 au total (Annexe 18). Une erreur inférieure à 1 signifie que l'erreur est inférieure d'un ordre de grandeur du seuil de perception à prédire. Au vu de l'incertitude de base sur les données de construction et la variabilité observable sur une partie des molécules comme vu au chapitre 2, la qualité de ce modèle est satisfaisante pour un modèle global.

La Figure 44 présente la répartition des groupements fonctionnels en fonction du logarithme décimal des erreurs de prédiction des seuils de perception par le modèle global.

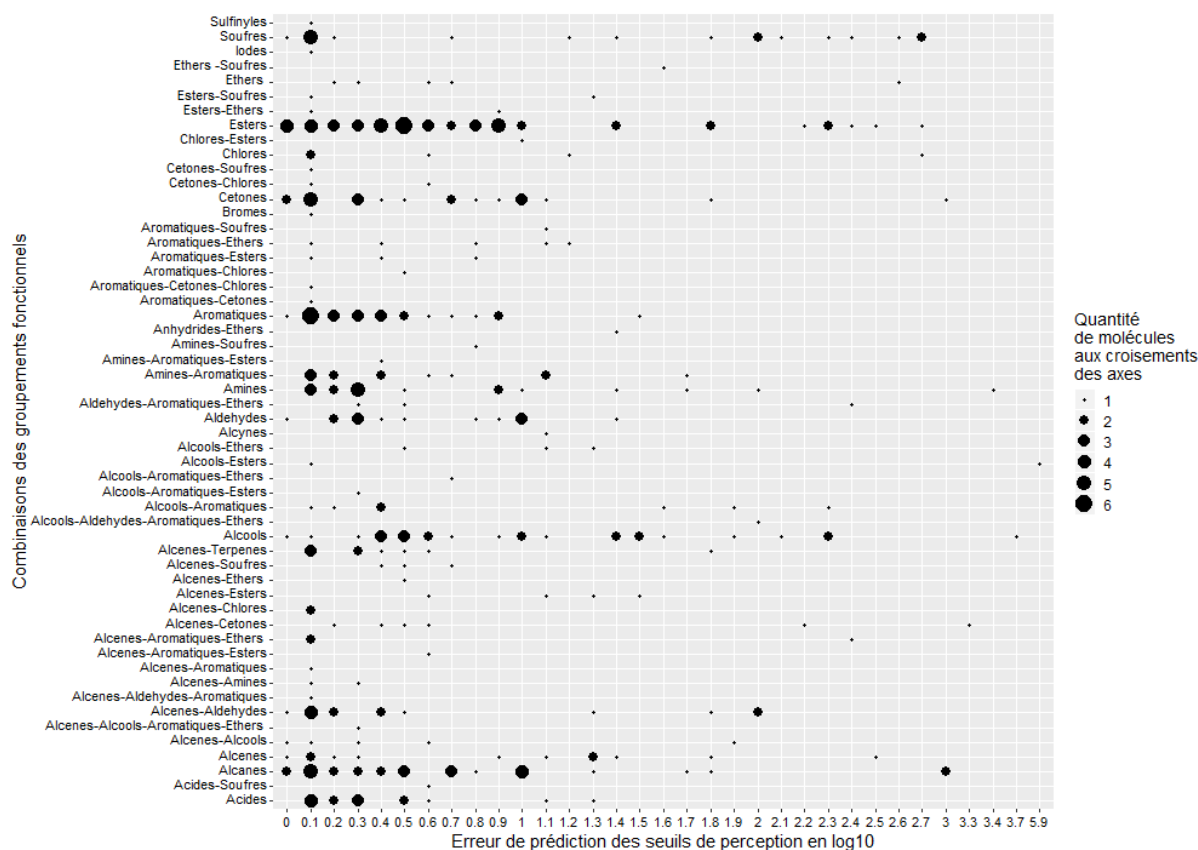


Figure 44 : Description des erreurs de prédiction des seuils de perception par le modèle global en fonction des groupements fonctionnels des molécules

On peut constater que les principales molécules dont le seuil de perception est prédit avec une erreur supérieure à 2 par le modèle global sont des molécules monofonctionnelles (principalement des composés soufrés, esters, alcools, alcanes et en moindre mesure alcènes, amines, chlorés, cétones et éthers). Six molécules polyfonctionnelles sont présentes dans cet espace :

- Erreur de 5,9 : Butyl Lactate (Butyl 2-hydroxypropanoate) (alcool-ester) ;
- Erreur de 3,3 : (E)-4-(2,6,6-triméthylcyclohex-2-en-1-yl)but-3-en-2-one (alcène-cétone)
- Erreur de 2,4 : 4,7-diméthoxy-5-prop-2-enyl-1,3-benzodioxole (alcène-aromatique-éther)
- Erreur de 2,4 : 3-(1,3-benzodioxol-5-yl)-2-méthylpropanal (aldéhyde-aromatique-éther)
- Erreur de 2,3 : 4-méthylphénol (alcool-aromatique)
- Erreur de 2,2 : 3,5,5-triméthylcyclohex-2-en-1-one (alcène-cétone)

Les configurations de groupements fonctionnels énumérés ci-dessus correspondent pourtant à des configurations retrouvées dans des molécules dont le seuil a été correctement prédit par le modèle global. Par exemple, la molécule dont le seuil est le moins bien prédit est une molécule bi fonctionnelle de type « alcool-ester ». Le seuil de perception de la seconde molécule bi-fonctionnelle de type « alcool-ester » présente dans la base de données, est quant à lui prédit avec exactitude : il s'agit du texanol ((3-hydroxy-

2,2,4-trimethylpentyl)-2-methylpropanoate) avec une erreur de 0,1. Ces deux molécules présentent les mêmes groupements fonctionnels en même proportion. Un autre facteur, non-identifié, semble donc être responsable de cette erreur.

Au vu de ces résultats, il n'apparaît pas de corrélation directe entre le nombre et la nature des groupements fonctionnels d'une molécule et la qualité de prédiction de son seuil de perception.

Ce constat encourage donc à poursuivre avec une modélisation locale.

3.6.2.2 Affinage du modèle global par modélisation locale des erreurs

Comme évoqué ci-dessus, affiner le modèle global nécessite donc de passer par une modélisation locale des erreurs. Pour cela la décomposition du jeu de données en catégories est nécessaire. La composition de chacune des catégories décrites au paragraphe 3.6.1.2 est présentée dans le Tableau 30.

Tableau 30 : Répartition des molécules par catégorie pour la modélisation locale de l'erreur

Catégories	Nombre de molécules	Proportion
Cat1 (0 ; 0,5 [)	182	52
Cat2 (0,5 ; 1[)	72	21
Cat3 (1 ; 1,5 [)	43	12
Cat4 ($\geq 1,5$)	53	15
	350	100

L'apport de la modélisation locale est présenté sous forme de graphe mettant en relation les valeurs prédites de seuils de perception exprimées en logarithme décimal (« Y Pred ») et les valeurs utilisées pour la prédiction (« Y Obs »). Les résultats sont présentés en Figure 45, Figure 46, Figure 47 et Figure 48 pour respectivement les catégories 1, 2, 3 et 4.

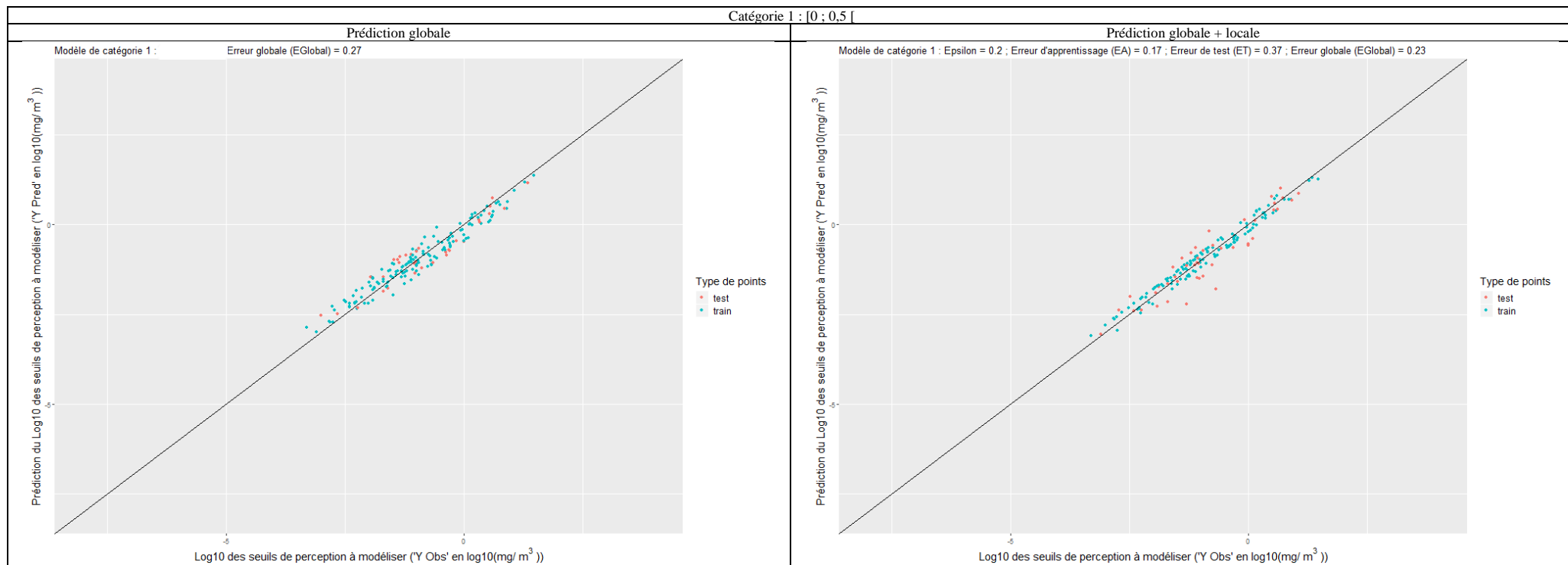


Figure 45 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 1, avant et après affinage par modélisation locale

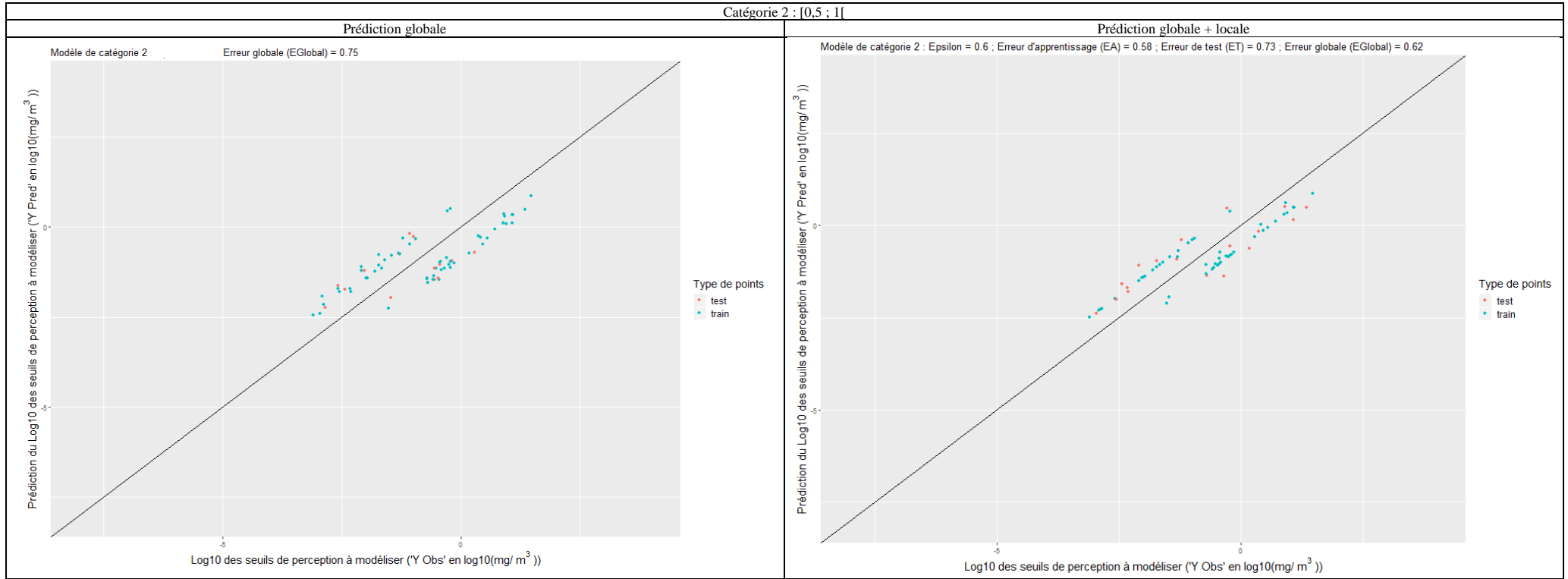


Figure 46 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 2, avant et après affinage par modélisation locale

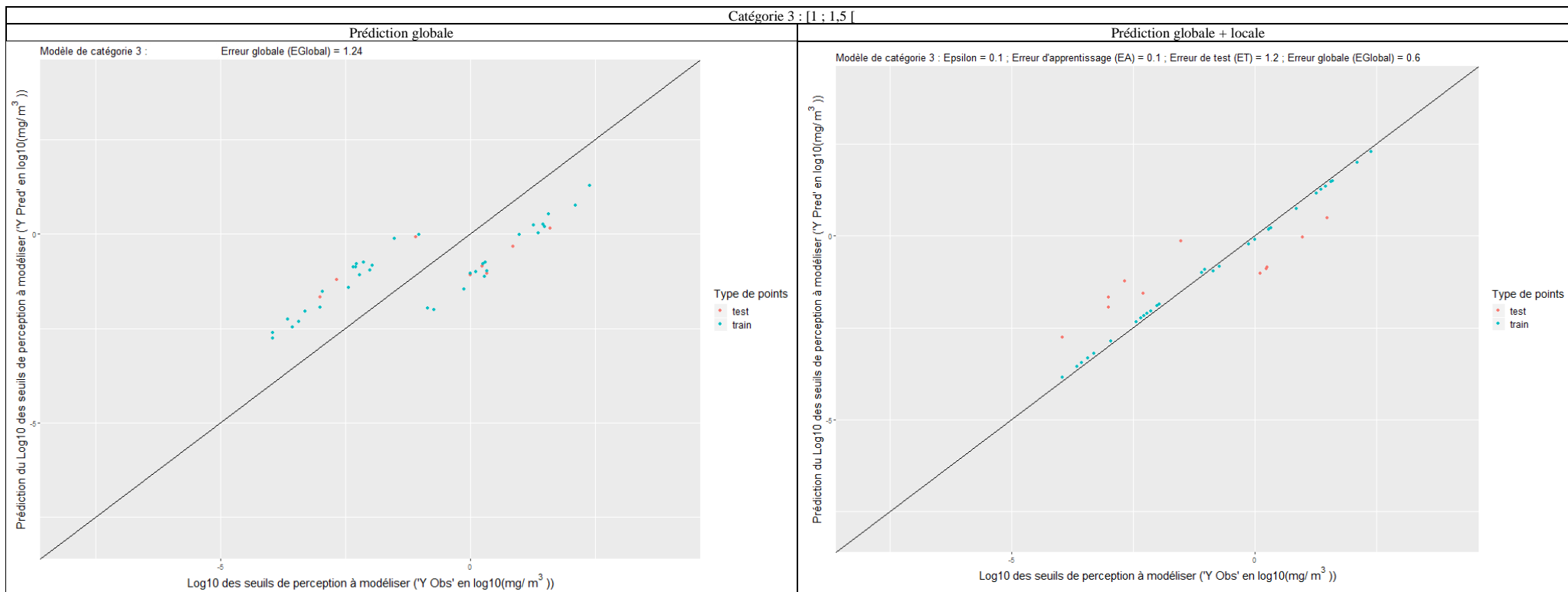


Figure 47 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 3, avant et après affinage par modélisation locale

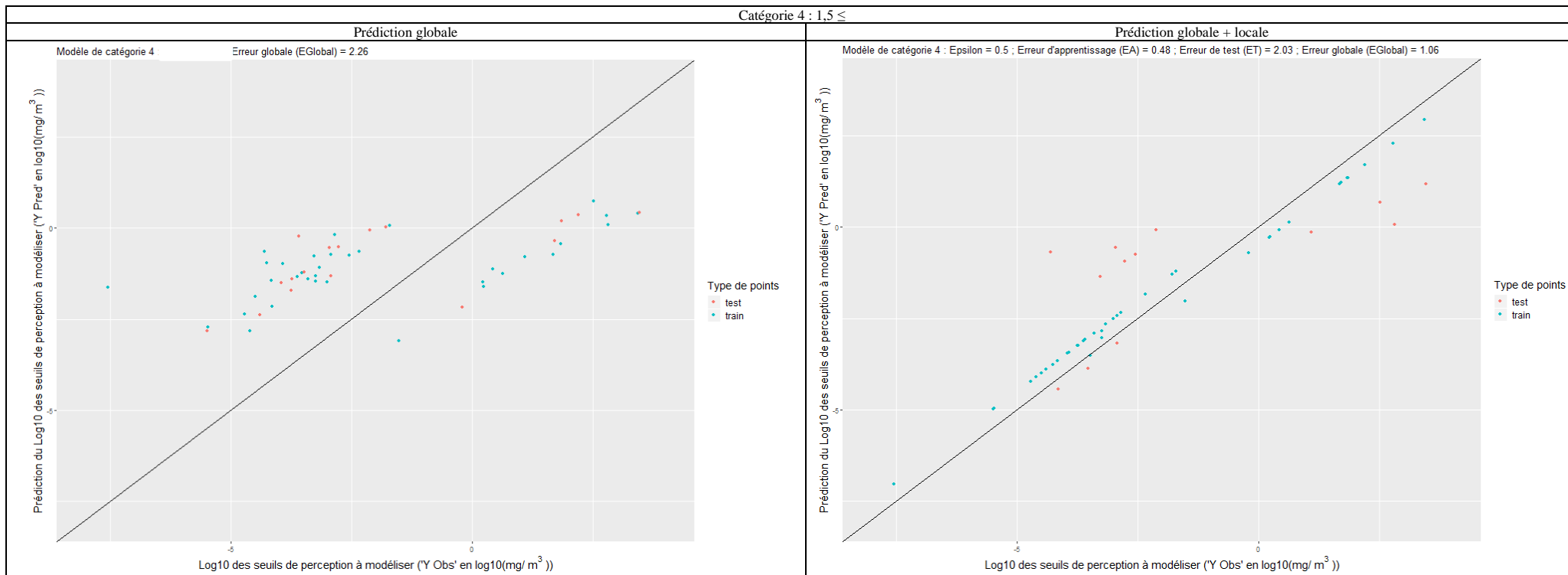


Figure 48 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 4, avant et après affinage par modélisation locale

Les graphes présentés ci-dessus montrent que l'affinage local du modèle global ne présente pas un intérêt important pour les catégories 1 et 2. En revanche, il montre de bonnes capacités pour les catégories 3 et 4 malgré un sur apprentissage très prononcé.

Les graphiques des seuils de perception prédits en fonction des seuils de perception observés sur l'ensemble du jeu de données de 350 molécules dans le cas du modèle global puis du modèle global affiné par modèles locaux sont présentés en Figure 49. Les données de ces modèles sont présentées en Annexe 20.

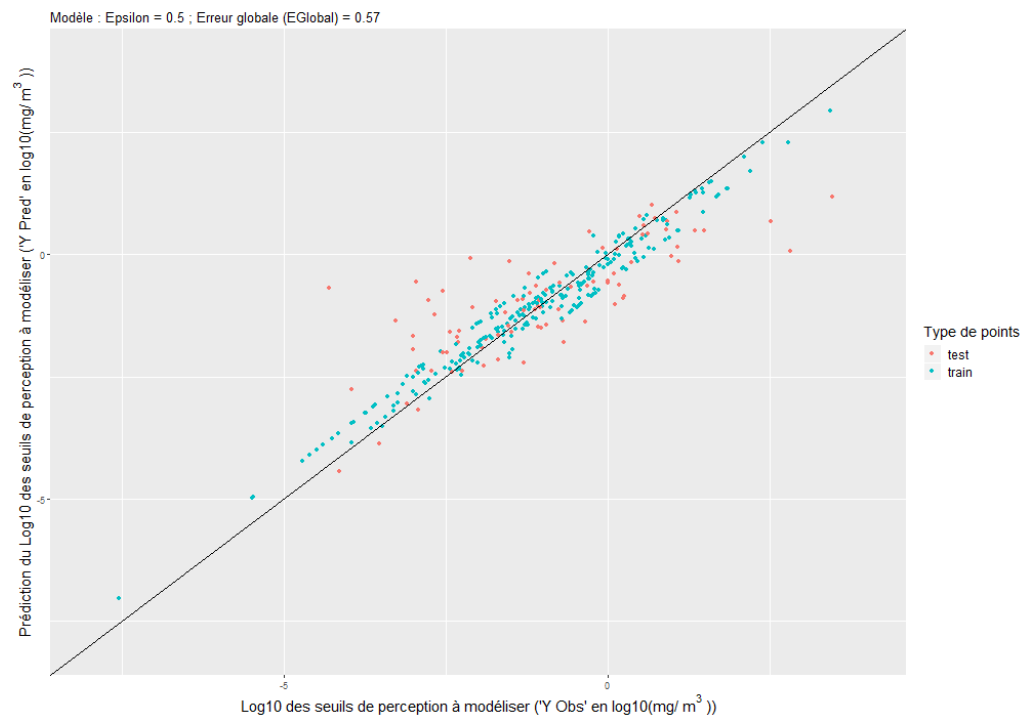
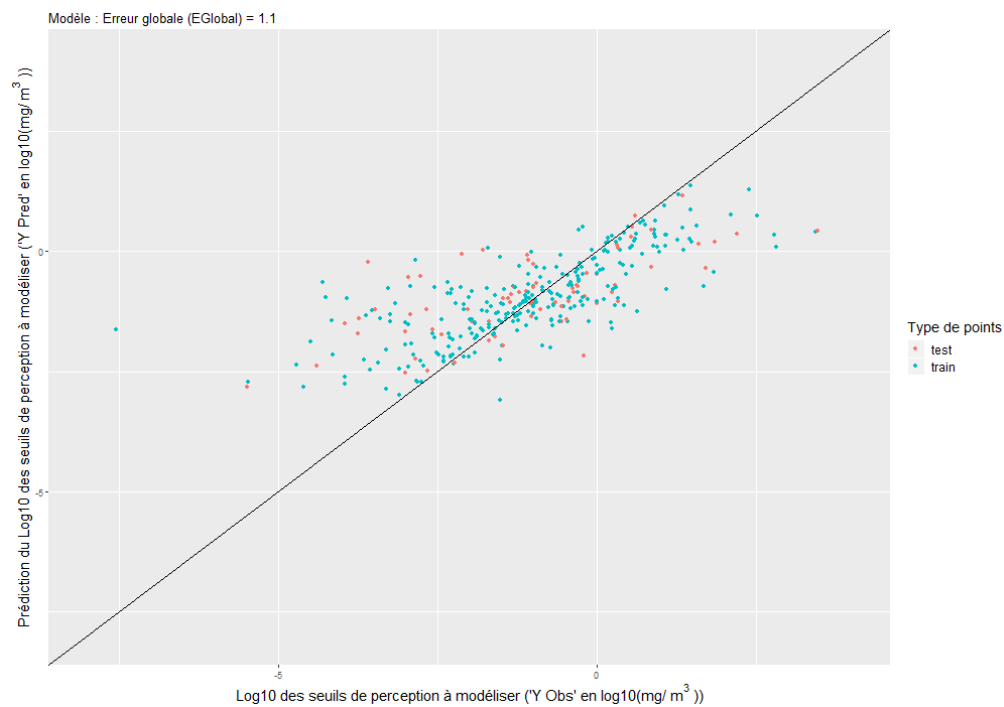


Figure 49 : Comparaison de la prédiction des seuils de perception des 350 molécules avant et après affinage par modélisation locale

L'affinage permet de diminuer l'erreur sur l'ensemble des données de 1,10 en échelle Log_{10} à 0,57.

Parmi les 350 molécules, il est choisi d'analyser le gain d'information sur les molécules pour lesquelles le seuil de perception a été mesuré par Olentica (Tableau 31).

Tableau 31 : Comparaison de la modélisation globale et de la modélisation globale additionnée de l'affinage sur les molécules dont le seuil de perception a été mesuré par Olentica

Famille	CAS	Molécules	Division du jeu de données pour le modèle global	Division du jeu de données pour les modèles locaux	Catégorie selon l'erreur du modèle global	SP mesuré par Olentica (mg/m^3)	SP prédit par le modèle global (mg/m^3)	SP affiné par modèles locaux (mg/m^3)
Aldéhydes	123-05-7	2-ethylhexanal	train	train	Cat1	0,0773	0,0284	0,0488
Esters	623-42-7	methyl butanoate	train	train	Cat1	0,0332	0,0789	0,0526
Cétones	108-83-8	2,6-dimethylheptan-4-one	train	train	Cat1	0,4009	0,1910	0,2530
Alcènes	110-83-8	cyclohexene	train	train	Cat1	0,2351	0,4563	0,3726
Alcools	104-76-7	2-ethylhexan-1-ol	train	train	Cat1	0,3809	0,2017	0,2403
Ethers	60-29-7	ethoxyethane	train	train	Cat1	0,1000	0,1875	0,1585
Aromatiques	60-12-8	2-phenylethanol	train	train	Cat1	0,0153	0,0241	0,0217
Azotés	121-44-8	N,N-diethylethanamine	train	train	Cat1	0,0401	0,0508	0,0636
Terpènes	99-85-4	1-methyl-4-propan-2-ylcyclohexa-1,4-diene	train	train	Cat1	0,4690	0,3726	0,5271
Aromatiques	100-42-5	styrene	train	train	Cat1	0,0592	0,0745	0,0757
Cétones	120-92-3	cyclopentanone	train	train	Cat1	0,9109	0,7242	1,0765
Esters	123-86-4	butyl acetate	train	train	Cat1	0,0513	0,0434	0,0364
Cétones	589-38-8	hexan-3-one	train	train	Cat1	0,1090	0,0979	0,1480
Alcènes	592-76-7	hept-1-ene	train	train	Cat1	0,1425	0,1463	0,2259
Aromatiques	108-88-3	toluene	train	test	Cat1	1,2166	0,4215	0,4077
Esters	140-11-4	benzyl acetate	train	test	Cat1	0,1729	0,0678	0,0734
Aldéhydes	66-25-1	hexanal	train	test	Cat1	0,0032	0,0070	0,0097
Cétones	110-43-0	heptan-2-one	train	test	Cat1	0,0781	0,1091	0,2227
Cétones	98-86-2	1-phenylethanone	train	test	Cat1	0,013	0,0176	0,0180
Soufrés	75-18-3	methylsulfanylmethane (DMS)	train	test	Cat1	0,0040	0,0050	0,0040
Esters	141-78-6	ethyl acetate	train	train	Cat2	2,9091	0,3311	0,7307
Cétones	108-10-1	4-methylpentan-2-one	train	train	Cat2	0,7212	0,1010	0,1894
Cétones	67-64-1	propan-2-one	train	train	Cat2	7,7548	1,2799	1,9477
Alcane	110-54-3	hexane	train	train	Cat2	12,4025	2,1457	3,1150
Cétones	78-93-3	butan-2-one	train	train	Cat2	2,5439	0,5151	1,0497
Ethers	1634-04-4	2-methoxy-2-methylpropane (MTBE)	train	train	Cat2	0,0250	0,1182	0,0995
Terpènes	5989-27-5	(4R)-1-methyl-4-prop-1-en-2-ylcyclohexene (D-Limonène)	train	train	Cat2	0,0829	0,3365	0,3301
Esters	624-24-8	methyl pentanoate	train	train	Cat2	0,0101	0,0371	0,0401
Alcools	71-36-3	butan-1-ol	train	train	Cat2	0,0521	0,1756	0,2070
Esters	105-37-3	ethyl propanoate	train	train	Cat2	0,0219	0,0697	0,0873
Aromatiques	95-47-6	1,2-xylene	train	test	Cat2	1,4936	0,1894	0,2380
Aldéhydes	112-31-2	decanal	train	test	Cat2	0,0028	0,0162	0,0100
Cétones	591-78-6	hexan-2-one	train	test	Cat2	0,6029	0,1105	0,2717
Ethers	109-99-9	oxolane (THF)	train	test	Cat2	2,2602	0,5601	0,6728
Alcools	106-21-8	3,7-dimethyloctan-1-ol	train	train	Cat3	0,0046	0,1325	0,0058
Aldéhydes	590-86-3	3-methylbutanal	train	train	Cat3	0,0002	0,0055	0,0003
Alcènes	111-66-0	oct-1-ene	train	train	Cat3	0,0072	0,1747	0,0090
Alcools	64-17-5	ethanol	train	train	Cat3	0,0936	0,9488	0,1178
Cétones	431-03-8	butane-2,3-dione	train	train	Cat4	0,0001	0,1067	0,0004
Ethers	470-82-6	2,2,4-trimethyl-3-oxabicyclo[2.2.2]octane	train	train	Cat4	0,0014	0,6418	0,0046
Soufrés	75-08-1	Ethanethiol	train	train	Cat4	0,00003	0,0132	0,0001
Esters	105-54-4	ethyl butanoate	train	train	Cat4	0,0006	0,0346	0,0014
Alcènes	592-41-6	hex-1-ene	train	train	Cat4	0,0005	0,1812	0,0007
Terpènes	80-56-8	4,6,6-trimethylbicyclo[3.1.1]hept-3-ene (alpha-pinène)	train	test	Cat4	0,0192	0,8791	0,7896

Comme cela a été précisé plus haut, toutes les molécules dont le seuil de perception a été mesuré par Olientica sont intégrées dans le jeu de données de construction du modèle global. Pour valider cette approche, il est nécessaire de vérifier qu'un gain d'information est apporté par la modélisation locale de l'erreur.

Pour ce qui concerne les molécules appartenant à la catégorie 1, le gain d'information obtenu est peu notable. Si la molécule appartient à la catégorie 2, la modélisation locale permet d'approcher la mesure d'Olientica ou d'obtenir une valeur similaire au modèle global y compris pour les molécules dont le seuil de perception n'est pas utilisé dans la construction du modèle local. Si la molécule appartient à la catégorie 3 ou à la catégorie 4, un gain d'information est observé sous réserve que la molécule appartienne au jeu de construction (« train ») du modèle local. En effet, le modèle local du hex-1-ene appartenant au jeu de validation de la catégorie 4 permet d'approcher la valeur initiale par rapport au modèle global mais de manière moins importante que pour les autres molécules de la même catégorie. Cela souligne un sur-apprentissage local. Cependant, il convient de noter au vu des graphes présentés plus haut et des résultats de l'Annexe 20 que le sur-apprentissage ne dégrade pas pour autant la valeur obtenue à l'aide du modèle global. Cela signifie que la valeur de prédiction obtenue par le modèle local est améliorée, sinon stable par rapport à celle obtenue par le modèle global.

3.7 Évaluation et limites de la modélisation

Un modèle local affiné localement est proposé. L'intérêt de cette modélisation est vérifié sur les molécules dont le seuil de perception est mesuré au laboratoire. Cependant ces données de seuils de perception sont toutes utilisées dans la construction du modèle global. La partie qui suit consiste à observer le résultat du modèle si aucune mesure de seuil de perception n'avait été réalisée par Olientica et si les molécules correspondantes n'avaient pas été forcées dans la construction du modèle. De plus la limite de prédiction du modèle sera soulevée.

3.7.1 Apport de l'utilisation des seuils de perception sans incertitude dans la conception du modèle

Afin d'évaluer la démarche autrement qu'avec un jeu de données de validation (« test »), le modèle est appliqué en considérant uniquement comme données d'entrée les seuils de perception obtenue après nettoyage des données de la littérature. Le jeu de données de construction (« train ») est alors constitué selon la démarche décrite en paragraphe 3.4.1 sans prise en compte des molécules dont le seuil de perception a été mesuré par Olientica. Néanmoins, une comparaison des seuils prédits dans ce cas (à partir des seuls seuils obtenus après nettoyage) est réalisée avec les seuils de perception mesurés par Olientica (Tableau 32). Cela permet d'observer trois aspects :

- L'amélioration ou non de la prédiction du seuil de perception de la littérature nettoyée grâce à la modélisation locale par rapport au modèle global ;
- La capacité ou non des modèles locaux à corriger une valeur d'entrée éloignée des mesures d'Olentica ;
- La qualité de la prédiction des molécules dont le seuil de perception n'est pas renseigné dans la littérature ou lorsque la molécule n'est ni dans le jeu de données de construction du modèle global ni dans celui du modèle local par rapport à la valeur expérimentale.

Pour ce dernier point, il est nécessaire de proposer un modèle local, donc une catégorie pour les molécules dont le seuil de perception est inconnu. Cela passe par une identification de la molécule renseignée ayant la structure la plus proche de la molécule inconnue. Pour cela, le jeu de données des molécules renseignées est décomposé à l'aide d'une analyse en composante indépendantes (ACI) de manière à décomposer ensuite le signal de la molécule non renseignée sur les mêmes composantes indépendantes. Cela permet de calculer le coefficient de corrélation entre les molécules renseignées et la molécule non renseignée. Le modèle local appliqué à la molécule non renseignée correspond alors au modèle local de la molécule renseignée ayant le coefficient de corrélation maximal avec cette molécule non renseignée.

Tableau 32 : Comparaison du résultat de la modélisation globale (avec et sans affinage local) avec les valeurs expérimentales d'Olentica (valeurs d'entrée du modèle = valeurs de la littérature nettoyée)

Famille	CAS	Molécules	Division du jeu de données pour le modèle global	Division du jeu de données pour les modèles locaux	Catégorie selon l'erreur du modèle global	SP de la littérature après nettoyage (mg/m^3)	SP prédit par le modèle global (mg/m^3)	SP affiné par modèles locaux (mg/m^3)	SP mesuré par Olentica (mg/m^3)
Alcènes	592-76-7	hept-1-ene	train	train	Cat1	1,5000	0,8834	0,9460	0,1425
Alcènes	592-41-6	hex-1-ene	train	train	Cat1	0,4800	0,7472	0,6213	0,0005
Cétones	110-43-0	heptan-2-one	train	train	Cat1	0,1280	0,1260	0,1925	0,0781
Cétones	67-64-1	propan-2-one	train	train	Cat1	2,9569	1,3102	4,6892	7,7548
Alcènes	110-83-8	cyclohexene	train	train	Cat1	0,6000	0,9152	0,9509	0,2351
Aldéhydes	112-31-2	decanal	train	train	Cat1	0,0364	0,0422	0,0577	0,0028
Ethers	60-29-7	ethoxyethane	train	train	Cat1	0,5000	0,3977	0,4486	0,1000
Soufrés	75-18-3	Methylsulfanylmethane(DMS)	train	train	Cat1	0,0015	0,0046	0,0023	0,0040
Aldéhydes	66-25-1	hexanal	train	train	Cat1	0,0247	0,0230	0,0263	0,0032
Esters	623-42-7	methyl butanoate	train	train	Cat1	0,0770	0,1792	0,0980	0,0332
Aromatiques	100-42-5	styrene	train	train	Cat1	0,1420	0,1126	0,1049	0,0592
Cétones	78-93-3	butan-2-one	train	train	Cat2	3,0651	0,7192	2,3960	2,5439
Cétones	591-78-6	hexan-2-one	train	train	Cat2	0,6787	0,1965	0,4226	0,6029
Aromatiques	60-12-8	2-phenylethanol	train	train	Cat2	0,0049	0,0182	0,0155	0,0153
Cétones	108-10-1	4-methylpentan-2-one	train	train	Cat2	0,7243	0,1578	0,2290	0,7212
Esters	141-78-6	ethyl acetate	train	train	Cat2	3,5735	0,4950	1,1297	2,9091
Terpènes	80-56-8	4,6,6-trimethylbicyclo[3.1.1]hept-3-ene (alpha-pinène)	train	train	Cat2	0,1517	1,2362	0,4800	0,0192
Aromatiques	95-47-6	1,2-xylene	train	train	Cat2	0,8902	0,1847	0,2810	1,4936
Cétones	431-03-8	butane-2,3-dione	train	train	Cat4	0,0019	0,1176	0,0024	0,0001
Aldéhydes	590-86-3	3-methylbutanal	train	train	Cat4	0,0004	0,0190	0,0004	0,0002
Soufrés	75-08-1	ethanethiol	train	train	Cat4	0,00003	0,0066	0,00004	0,00003
Esters	105-54-4	ethyl butanoate	train	train	Cat4	0,00005	0,0637	0,0001	0,0006
Alcools	71-36-3	butan-1-ol	test	train	Cat1	0,1638	0,3293	0,2598	0,0521
Esters	123-86-4	butyl acetate	test	train	Cat1	0,0525	0,1054	0,0833	0,0513
Alcools	64-17-5	ethanol	test	train	Cat1	1,1080	0,7534	0,9455	0,0936
Esters	105-37-3	ethyl propanoate	test	train	Cat1	0,1942	0,1523	0,3079	0,0219
Alcane	110-54-3	hexane	test	train	Cat1	5,3000	3,3420	3,3400	12,4025
Aromatiques	108-88-3	toluene	test	train	Cat1	1,1326	0,4401	1,0747	1,2166
Azotés	121-44-8	N,N-diethylethanamine	test	train	Cat1	0,1483	0,1296	0,2351	0,0401
Terpènes	99-85-4	1-methyl-4-propan-2-ylcyclohexa-1,4-diene	test	train	Cat2	2,5000	0,3379	0,7903	0,4690

Esters	624-24-8	methyl pentanoate	test	test	Cat2	0,0110	0,0953	0,0500	0,0101
Alcènes	111-66-0	oct-1-ene	test	train	Cat4	0,0046	1,1156	0,0058	0,0072
Ethers	470-82-6	2,2,4-trimethyl-3-oxabicyclo[2.2.2]octane	test	train	Cat4	0,0313	1,3137	0,0394	0,0014
Cétones	98-86-2	1-phenylethanone	NA	NA	Cat1	NA	0,0213	0,0193	0,0130
Cétones	120-92-3	cyclopentanone	NA	NA	Cat1	NA	0,5964	0,9975	0,9109
Ethers	1634-04-4	2-methoxy-2-methylpropane (MTBE)	NA	NA	Cat1	NA	0,3123	0,4356	0,0250
Terpènes	5989-27-5	(4R)-1-methyl-4-prop-1-en-2-ylcyclohexene (D-Limonène)	NA	NA	Cat1	NA	0,4542	0,5313	0,0829
Ethers	109-99-9	oxolane (THF)	NA	NA	Cat1	NA	0,4844	1,0043	2,2602
Cétones	108-83-8	2,6-dimethylheptan-4-one	NA	NA	Cat2	NA	0,0462	0,0022	0,4009
Aldéhydes	123-05-7	2-ethylhexanal	NA	NA	Cat2	NA	0,0141	0,0086	0,0773
Aldéhydes	96-17-3	2-methylbutanal	NA	NA	Cat2	NA	0,0184	0,0064	0,0017
Alcools	106-21-8	3,7-dimethyloctan-1-ol	NA	NA	Cat2	NA	0,4234	0,0065	0,0046
Cétones	589-38-8	hexan-3-one	NA	NA	Cat2	NA	0,1477	1,2055	0,1090
Esters	140-11-4	benzyl acetate	NA	NA	Cat2	NA	0,0392	0,1721	0,1729
Alcools	104-76-7	2-ethylhexan-1-ol	NA	NA	Cat3	NA	0,2624	0,5054	0,3809

Tout d'abord, un point est fait sur les molécules ayant servi à la construction à la fois du modèle global et du modèle local (« train ») :

- Pour les molécules de la catégorie 1 (« Cat1 ») on constate que dans 5 cas sur 11 la valeur prédite par le modèle local est plus proche de la valeur d'entrée que celle obtenue par le modèle global. La situation inverse se retrouve dans 2 cas sur 11. Pour les 4 cas restants, on ne constate ni amélioration ni dégradation de la prédiction par le modèle local. Ces résultats illustrent l'intérêt d'affiner le modèle global par des modèles locaux. En revanche, quand une trop grande différence existe entre le seuil de perception nettoyé et celui mesuré par Olentica (hex-1-ène), l'affinage du seuil de perception ne conduit pas à un rapprochement entre les deux valeurs. Le modèle ne permet donc pas de pallier cette différence et nécessite d'être construit sur des données fiables.
- Pour les molécules de la catégorie 2 et de la catégorie 4, une amélioration de la prédiction de la valeur d'entrée par la modélisation locale est observée pour l'ensemble des molécules. Néanmoins, la modélisation locale ne permet à nouveau pas d'approcher la valeur expérimentale si celle-ci est éloignée de la valeur d'entrée comme c'est le cas pour le 2-phényléthanol ou l'alpha-pinène (4,6,6-trimethylbicyclo[3.1.1]hept-3-ene).
- Le cas n'est pas illustré pour la catégorie 3.

Concernant les molécules incluses dans le jeu de données de validation du modèle global (« test ») et présentes dans le jeu de données de construction des modèles locaux (« train ») :

- Pour les molécules des catégories 1, 2 et 4, le modèle local permet d'améliorer ou de stabiliser la prédiction des valeurs d'entrée par rapport au modèle global mais ne permet pas d'approcher la valeur mesurée si celle-ci est éloignée de la valeur d'entrée. L'amélioration augmente avec l'erreur de la catégorie.
- Le cas n'est pas illustré pour la catégorie 3.

Enfin, en considérant les 13 molécules pour lesquelles aucune information n'est disponible initialement (« NA » dans le tableau) ou dont la valeur du seuil de perception est utilisée dans le jeu de validation

pour à la fois le modèle global et le modèle local, le nombre de scénarios positifs est supérieur au nombre de scénarios négatifs. En effet, dans 10 cas sur 13, la modélisation locale permet soit d'approcher la valeur expérimentale par rapport au modèle global soit à minima de la stabiliser. Les 3 molécules pour lesquelles la modélisation locale dégrade le résultat par rapport au modèle global sont la 2,6-diméthylheptan-4-one, la 2-éthylhexanal et la hexan-3-one.

En conclusion, l'utilisation d'un modèle local permet d'améliorer et à minima de stabiliser la prédiction du modèle global dans la plupart des cas. Néanmoins, il ne permet pas de corriger la valeur d'entrée (seuil de perception de la littérature nettoyé) si celle-ci est éloignée de la valeur mesurée par Olentica. Lorsque les molécules n'ont pas de seuils de perception renseignés dans la littérature ou lorsque celle-ci n'est jamais utilisée dans un jeu de construction de modèle, le modèle local permet d'approcher le seuil de perception mesurée par Olentica dans 8 cas sur 13 (cas du 3,7-dimethyloctan-1-ol). Puis dans 2 cas sur 13, le modèle local maintient la prédiction du modèle global sans l'améliorer comme cela est le cas pour D-Limonène ((4R)-1-methyl-4-prop-1-en-2-ylcyclohexene). Elle induit dans 3 cas sur 13 un éloignement (cas de la hexan-3-one). L'affinage par un modèle local permet de prédire un seuil de perception selon une double modélisation qui renforce la valeur de prédiction. Ainsi, même si, a priori, la modélisation locale n'apporte pas un gain important dans le cas des catégories 1 et 2, elle permet néanmoins d'approcher la valeur expérimentale pour les molécules dont le seuil de perception n'est pas renseigné dans le modèle pour ces catégories. Pour les catégories 3 et 4, la modélisation locale montre également de bonne performance malgré les inconvénients du sur-apprentissage. L'une des raisons de ce sur apprentissage local peut être liée à la phase de division du jeu de données en jeu de données de construction (« train ») et en jeu de données de validation (« test ») puisque c'est une phase ne dépendant que de la représentativité des seuils de perception présents et non de leur structure. Choisir un recouvrement adéquat comme appliqué pour le modèle global est une piste d'amélioration. Une deuxième piste d'amélioration serait également de trouver un autre critère de sélection du nombre de composantes indépendantes ACI.

Ces résultats sont donc encourageants par la capacité qu'a le modèle global affiné par les modèles locaux d'approcher correctement des valeurs inconnues (comme observé pour les 13 molécules mesurées expérimentalement mais non utilisées pour la construction du modèle ou encore pour le méthyl pentanoate qui a servi deux fois dans le « test »).

3.7.2 Limite d'application du modèle

La démarche de modélisation a été mise en place pour prédire le seuil de perception des molécules non renseignées dans la littérature. Cette modélisation se base sur la structure des molécules dont le seuil de perception est disponible. La structure de ces dernières molécules a permis de définir un espace de

représentativité du modèle à l'aide de deux axes obtenus par la méthode de « Projection Pursuit ». En projetant les molécules dont le seuil de perception est inconnu dans cet espace il est possible de juger de la pertinence du seuil modélisé pour ces molécules. Ces dernières correspondent à des molécules identifiées dans diverses études menées au laboratoire et sont inventoriées au nombre de 893. Après nettoyage du chapitre 2, ce nombre est de 1193.

Pour évaluer la pertinence de notre modèle sur ces 1193 molécules, le traitement statistique suivant a été mis en place. Celui-ci consiste à reproduire exactement les mêmes étapes de calcul que ceux appliqués aux molécules utilisées pour le modèle :

Soit X la matrice initiale de variables d'indices à décomposer. Celle-ci est tout d'abord centrée-réduite et les attributs de centrage et standardisation sont récupérés sous respectivement les termes «COLMEAN » et « COLSD ». Cette matrice centrée-réduite obtenue est notée $X1$. $X1$ subit un « blanchiment » à l'aide d'une matrice de blanchiment K permettant de projeter la matrice $X1$ sous ces directions principales et préalablement déterminées pour obtenir $X2$. Puis la matrice $X2$ subit une dernière transformation par une matrice de « non-mixage » W qui maximise le caractère non gaussien des sources (MARCHINI et al., 2017b). Cela permet d'obtenir la matrice S contenant les coordonnées des molécules renseignées sur les axes de « Projection Pursuit ».

Cette transformation est appliquée sur la matrice contenant les indices des molécules dont le seuil de perception n'est pas renseigné (X_{proj}) et permet d'obtenir la matrice contenant les coordonnées des molécules non renseignées sur les axes de « Projection Pursuit » (S_{proj}). Il est alors obtenu toutes les molécules sous le même référentiel. Les deux jeux de coordonnées sont projetés sur un même graphique (Figure 50).

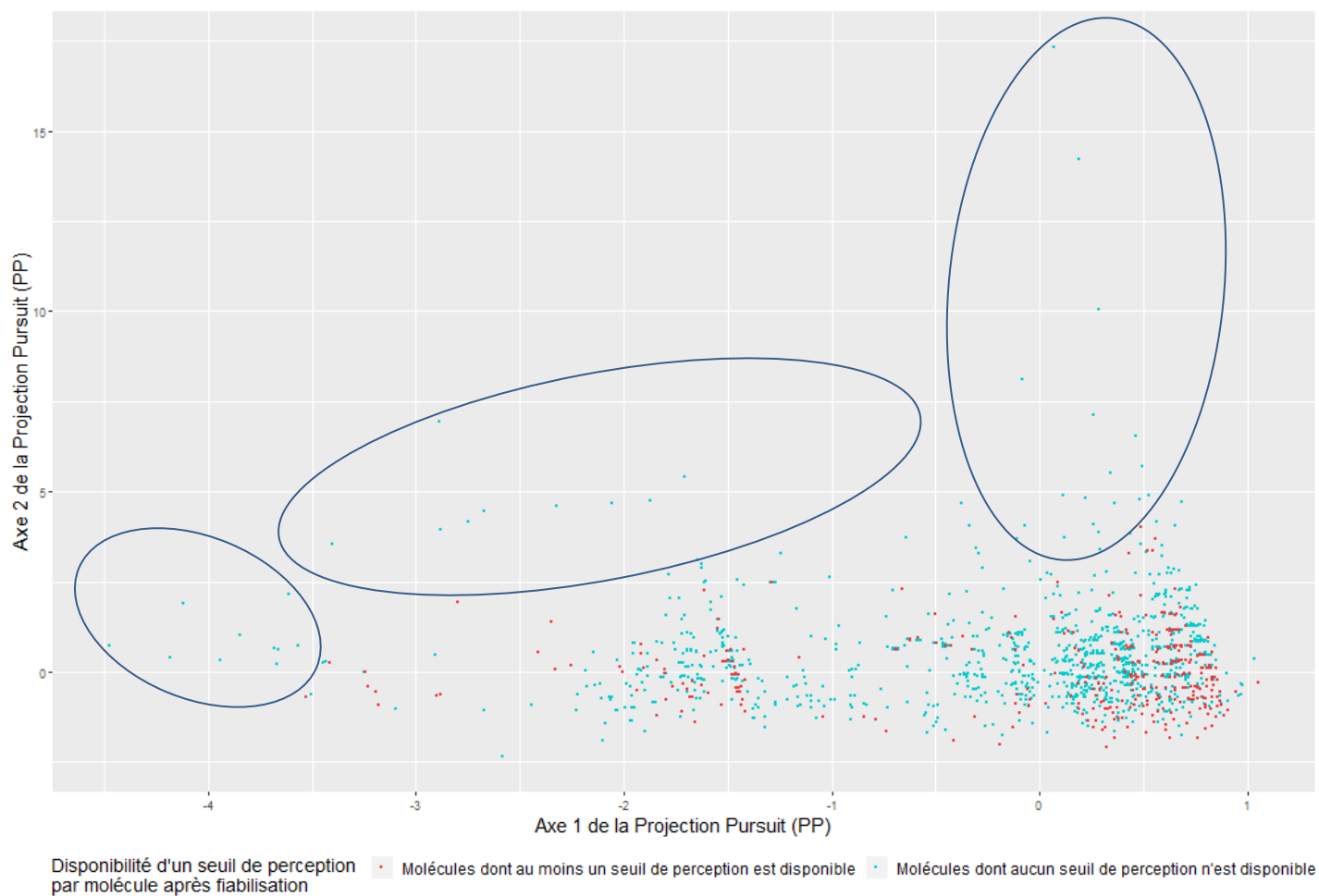


Figure 50 : Projection des molécules dont le seuil de perception n'est pas renseigné dans la littérature dans l'espace défini par les axes principaux de « Projection Pursuit » obtenus à partir des molécules ayant au moins un seuil de perception renseigné

La Figure 50 permet de constater que plusieurs molécules dont le seuil de perception n'est pas disponible sont éloignées de l'espace défini par les molécules ayant un seuil de perception disponible. Il s'agit des « zones d'ombres », indiquée par les cercles bleus en Figure 50. Quelle que soit la couverture et la fiabilité du modèle mis en place plus haut, celui-ci ne permettra pas une prédiction fiable des molécules dans ces zones.

Autrement dit, le modèle mis en place dans cette étude permet de prédire le seuil de perception des molécules présentes dans ces « zones d'ombres » mais avec une qualité moindre que les molécules se situant dans des zones fortement renseignées. Les molécules concernées sont au nombre de 33 environ sur les 1193 testées et sont présentées en Tableau 33 avec leurs coordonnées.

Tableau 33 : Molécules situées hors de la couverture de recouvrement des molécules renseignées

CAS	Axe 1 de la « Projection Pursuit »	Axe 2 de la « Projection Pursuit »	Molécules
112-80-1	-0,08	8,12	(Z)-octadec-9-enoic acid
134-81-6	-4,12	1,89	1,2-diphenylethane-1,2-dione
84-69-5	-1,71	5,41	bis(2-methylpropyl) benzene-1,2-dicarboxylate
629-73-2	0,11	4,90	hexadec-1-ene
109-43-3	0,29	10,05	dibutyl decanedioate
112-66-3	0,48	4,78	dodecyl acetate
1153-51-1	-1,87	4,74	(3R,5S,8R,9S,10S,13R,14S)-10,13-dimethyl-2,3,4,5,6,7,8,9,11,12,14,15-dodecahydro-1H-cyclopenta[a]phenanthren-3-ol
7148-51-8	-1,87	4,74	(3S,5S,8R,9S,10S,13R,14S)-10,13-dimethyl-2,3,4,5,6,7,8,9,11,12,14,15-dodecahydro-1H-cyclopenta[a]phenanthren-3-ol
4390-04-9	0,68	4,70	2,2,4,4,6,8,8-heptamethylnonane
544-63-8	0,36	4,68	tetradecanoic acid
3661-77-6	-0,37	4,68	cycloheptadecanone
18339-16-7	-2,06	4,67	(5S,8R,9S,10S,13R,14S)-10,13-dimethyl-1,2,4,5,6,7,8,9,11,12,14,15-dodecahydrocyclopenta[a]phenanthren-3-one
4075-07-4	-2,32	4,59	(8S,9S,10R,13R,14S)-10,13-dimethyl-1,2,6,7,8,9,11,12,14,15-decahydrocyclopenta[a]phenanthren-3-one
560-88-3	-2,67	4,44	(4,7,7-trimethyl-3-bicyclo[2.2.1]heptanyl) 2-hydroxybenzoate
21145-77-7	-2,75	4,18	1-(3,5,5,6,8,8-hexamethyl-6,7-dihydronaphthalen-2-yl)ethanone
84-61-7	-2,89	6,94	dicyclohexyl benzene-1,2-dicarboxylate
55282-12-7	0,19	14,21	3-ethyl-5-(2-ethylbutyl)octadecane
6846-50-0	0,26	7,12	[2,2,4-trimethyl-3-(2-methylpropanoyloxy)pentyl] 2-methylpropanoate
85-01-8	-4,47	0,74	Phenanthrene
86-73-7	-4,18	0,39	9H-Fluorene
2765-11-9	0,22	4,81	pentadecanal
122-39-4	-3,94	0,34	N-phenylaniline
119-61-9	-3,84	1,02	diphenylmethanone
644-08-6	-3,68	0,67	1-methyl-4-phenylbenzene
92-52-4	-3,67	0,23	1,1'-biphenyl
101-84-8	-3,66	0,61	phenoxybenzene
120-51-4	-3,61	2,14	benzyl benzoate
101-81-5	-3,57	0,73	benzylbenzene
593-49-7	0,07	17,33	heptacosane
544-76-3	0,53	4,90	hexadecane
544-76-3	0,53	4,90	hexadecane
629-78-7	0,50	5,68	heptadecane
4860-03-1	0,34	5,52	1-chlorohexadecane
593-45-3	0,46	6,53	octadecane

L'une des pistes d'amélioration pour pallier à ce manque de recouvrement de l'espace par les molécules dont le seuil de perception est renseigné, serait de mesurer le seuil de perception des molécules les plus en marge afin d'agrandir cet espace.

3.8 Conclusion sur la complétude des seuils de perception manquants

L'objectif de ce chapitre était de proposer une méthodologie permettant de prédire les seuils de perception de molécules considérées comme non fiables ou non disponibles dans la littérature. Cette modélisation se base sur un jeu de données consolidées de 350 molécules correspondant aux 45 seuils de perception mesurés par Olentica additionnés des seuils de la littérature validés suite à la mise en œuvre de la procédure de nettoyage de la base de données décrite dans le chapitre 2.

Ces données d'entrée ont permis d'obtenir un modèle global satisfaisant sur la base des critères suivants :

- une erreur de 1,06 sur le jeu de données de construction et 1,23 sur le jeu de données de validation exprimé en logarithme décimal (\log_{10}).
- un coefficient de corrélation de 0,76 pour le jeu de donnée de construction et de 0,70 pour le jeu de données de validation.

Pour améliorer ce modèle global, un affinage par des modèles locaux par catégorie d'erreur a été appliqué. En effet, ces erreurs sous-tendent une structure locale commune qui ne serait pas identifiée par la première approche globale. Pour cela, l'erreur de prédiction du seuil de perception de chaque molécule a été recensée pour permettre une catégorisation de ces molécules en fonction de cette erreur. Cette démarche permet de prédire le seuil de perception des molécules selon un double niveau d'échelle. Quatre catégories ont été établies.

Un gain d'information grâce à la modélisation locale a été observé pour les molécules dont le seuil de perception a été mesuré lors de notre étude.

- Catégorie 1 : pas de gain d'information utile mais de bons résultats pour la prédiction des valeurs inconnues (constaté pour 3 valeurs sur 5).
- Catégorie 2 : un léger gain d'information mais un apport controversé pour les valeurs inconnues avec dans 3 cas sur 7 un éloignement par rapport à la valeur mesurée par Olentica.
- Catégories 3 et 4 : un très bon gain d'information pour ces deux catégories malgré un sur-apprentissage qui ne dégrade néanmoins pas la valeur obtenue par le modèle global. De plus, sur l'unique valeur inconnue de catégorie 3, une bonne prédiction a été constatée.

L'affinage permet par conséquent, de diminuer l'erreur, exprimée en logarithme décimal, de 1,10 à 0,57 sur l'ensemble des données. Cela apparaît comme très encourageant.

En conclusion, le modèle affiné, même s'il est clairement perfectible, permet d'obtenir une valeur de seuils de perception correcte pour les molécules situées dans l'espace de construction. Celui-ci est d'autant plus utile que la catégorie d'erreur est grande.

En effet, il permet fortement d'améliorer les prédictions du modèle global pour les catégories 3 et 4 mais est moins pertinent dans le cas des catégories 1 et 2 où une légère amélioration sinon une stabilisation du résultat du modèle global est constatée. Les modèles locaux ne dégradent que très rarement la valeur utilisée pour la modélisation. On note cependant un sur apprentissage marqué.

Lorsque la valeur de seuil de perception est inconnue, le modèle affiné montre de bonnes capacités de prédiction des valeurs inconnues cohérente avec les mesures Olentica dans 10 cas sur 13.

Afin de limiter le sur-apprentissage de la phase d'affinage, une des pistes d'amélioration serait de proposer, au même titre que pour le modèle global, une constitution du jeu de construction couvrant à la fois l'espace des structures des molécules et la représentativité des seuils de perception, et non pas uniquement cette dernière.

Une autre piste d'amélioration serait d'attribuer dès le début de la démarche (au chapitre 2) un poids à chacun des articles pour faire une moyenne en fonction de leur « fréquence d'exclusion » (FE) et de leur « taux de comparaison » (TC) afin de conserver toute l'information initiale. Cela permettrait de pondérer l'information à l'aide de fonctions de croyance pour mieux représenter l'incertitude de base tout en la conservant et donc mieux gérer le sur-apprentissage (P. XU, DAVOINE, et DENOEUUX 2014).

Enfin, une amélioration des modèles locaux passerait par une meilleure sélection du nombre de composantes indépendantes en ACI.

CHAPITRE 4 : MISE AU POINT D'UNE DEMARCHE AUTOMATISEE D'IDENTIFICATION DES MOLECULES RESPONSABLES DE L'ODEUR D'UNE MATRICE GAZEUSE

4.1 Contexte

Si les chapitres précédents proposent une manière de gérer une partie de la variabilité liée aux données (celles de seuils de perception), l'objectif de ce chapitre est de proposer une manière de gérer une partie de la variabilité et des inconvénients liés à l'expert lui-même, cités au chapitre 1.

En effet, ce chapitre va présenter une démarche automatisée permettant d'identifier les molécules responsables de l'odeur d'un matériau et plus spécifiquement de la concentration d'odeur tout en palliant l'hétérogénéité du résultat de l'expert liée à la disparité des compétences, à la finesse d'analyse et au temps consacré. En effet, cette démarche automatisée a aussi pour but de réduire l'aspect chronophage de la démarche de l'expert. Enfin, cette démarche permettra également à l'industriel d'avoir une certaine indépendance sur l'analyse de ses résultats.

La démarche de l'expert se caractérise par l'utilisation des activités d'odeur (Équation 3 du chapitre 1) des molécules pour déterminer celles qui possèdent un poids important dans la concentration de l'odeur du mélange gazeux.

Ainsi, la qualité de l'identification dépendra de la précision du calcul de l'activité d'odeur des molécules et donc de deux types de données :

- Les données physico-chimiques caractérisant et quantifiant les molécules de la matrice gazeuse ;
- Les données de seuils de perception.

Afin de pallier au mieux l'incertitude liée à ces deux types de données et mettre au point notre démarche automatisée de la manière la plus réaliste possible, nous avons choisi de la mettre en place en utilisant des mélanges gazeux synthétiques. Cela permettra de travailler avec des mélanges dont la nature et la concentration des molécules sont connues et le seuil de perception de 75% d'entre elles validé par une mesure réalisée par Olentica (cf partie 2.7.1.1). Cette maîtrise de la génération des mélanges permettra de faire varier leur composition pour en modifier les caractéristiques olfactives. De cette façon, il sera aisé pour l'expert d'identifier les molécules responsables de l'odeur et les paramètres de la démarche automatisée seront choisis de manière à faire converger ses résultats avec ceux de l'expert sur ces mélanges.

On trouvera dans ce chapitre une partie méthodologique portant sur la fabrication des mélanges gazeux synthétiques et sur la démarche automatisée mise en place, et une partie présentant les résultats obtenus par cette approche sur les mélanges gazeux expérimentaux de référence.

4.2 Matériels et méthodes

Pour mettre au point la démarche automatisée de mise en relation des données physico-chimiques et sensorielles, la première étape est la génération de plusieurs mélanges gazeux synthétiques qui constitueront la base de données utilisée dans ce travail.

4.2.1 Fabrication des mélanges gazeux synthétiques

La constitution des mélanges gazeux synthétiques comporte deux étapes qui sont la préparation d'un mélange de molécules à l'état liquide puis la volatilisation de celui-ci. Ces deux phases sont décrites ci-dessous.

4.2.1.1 Préparation des mélanges liquides synthétiques à volatiliser

La première étape pour fabriquer un mélange gazeux synthétique est la préparation d'un mélange liquide. La constitution de ces mélanges liquides a été régie par plusieurs critères :

- Disposer de molécules présentant des seuils de perception d'ordres de grandeur différents ;
- Être en capacité de volatiliser le mélange liquide avec une température maximum de 250°C sans génération de produits de réaction ;
- Vérifier facilement, par analyse TD-GC-MS, l'absence de sous-produits et/ou de produits de réaction.

La réactivité des molécules type acides et amines primaires et secondaires fait qu'elles ne sont pas utilisées dans notre étude. En conséquence, les mélanges seront composés à partir des familles suivantes : les alcanes, les alcènes, les aromatiques, les alcools, les aldéhydes, les cétones, les esters et les éthers.

Le Tableau 34 présente les molécules sélectionnées qui composent les mélanges synthétiques ainsi que leurs seuils de perception.

Tableau 34 : Molécules et seuils de perception utilisés pour la constitution des 5 mélanges gazeux synthétiques

Famille	No_CAS	Molécules	Seuils de perception (mg/m^3)	Source
Alcanes	109-66-0	pentane	4,10E+00	Littérature
Alcanes	110-54-3	hexane	1,24E+01	Mesure expérimentale
Alcanes	111-65-9	octane	1,22E+01	Mesure expérimentale
Alcanes	111-84-2	nonane	1,20E+01	Littérature
Alcanes	1120-21-4	undecane	9,59E+00	Littérature
Alcanes	124-18-5	decane	1,60E+02	Littérature
Alcanes	142-82-5	heptane	5,71E+01	Littérature
Alcanes	75-83-2	2,2-dimethylbutane	7,00E+01	Littérature
Alcanes	79-29-8	2,3-dimethylbutane	1,50E+00	Littérature
Alcanes	96-14-0	3-methylpentane	3,10E+01	Littérature
Alcènes	110-83-8	cyclohexene	2,35E-01	Mesure expérimentale

Alcènes	111-66-0	oct-1-ene	7,18E-03	Mesure expérimentale
Alcènes	592-41-6	hex-1-ene	5,22E-04	Mesure expérimentale
Alcènes	592-76-7	hept-1-ene	1,43E-01	Mesure expérimentale
Alcools	104-76-7	2-ethylhexan-1-ol	3,81E-01	Mesure expérimentale
Alcools	64-17-5	ethanol	9,36E-02	Mesure expérimentale
Alcools	71-36-3	butan-1-ol	5,21E-02	Mesure expérimentale
Aldéhydes	112-31-2	decanal	2,81E-03	Mesure expérimentale
Aldéhydes	123-05-7	2-ethylhexanal	7,73E-02	Mesure expérimentale
Aldéhydes	124-13-0	octanal	5,00E-04	Mesure expérimentale
Aldéhydes	590-86-3	3-methylbutanal	2,20E-04	Mesure expérimentale
Aldéhydes	96-17-3	2-methylbutanal	1,68E-03	Mesure expérimentale
Aromatiques	100-41-4	ethylbenzene	3,16E-01	Littérature
Aromatiques	100-42-5	styrene	5,92E-02	Mesure expérimentale
Aromatiques	106-42-3	1,4-xylene	5,79E-01	Littérature
Aromatiques	108-38-3	1,3-xylene	5,08E-01	Littérature
Aromatiques	108-88-3	toluene	1,22E+00	Mesure expérimentale
Aromatiques	95-47-6	1,2-xylene	1,49E+00	Mesure expérimentale
Aromatiques	95-63-6	1,2,4-trimethylbenzene	6,43E-01	Littérature
Cétones	108-10-1	4-methylpentan-2-one	7,21E-01	Mesure expérimentale
Cétones	108-83-8	2,6-dimethylheptan-4-one	4,01E-01	Mesure expérimentale
Cétones	110-43-0	heptan-2-one	7,81E-02	Mesure expérimentale
Cétones	431-03-8	butane-2,3-dione	1,17E-04	Mesure expérimentale
Cétones	589-38-8	hexan-3-one	1,09E-01	Mesure expérimentale
Cétones	591-78-6	hexan-2-one	6,03E-01	Mesure expérimentale
Cétones	67-64-1	propan-2-one	7,75E+00	Mesure expérimentale
Cétones	78-93-3	butan-2-one	2,54E+00	Mesure expérimentale
Cétones	98-86-2	1-phenylethanone	1,30E-02	Mesure expérimentale
Esters	105-37-3	ethyl propanoate	2,19E-02	Mesure expérimentale
Esters	105-54-4	ethyl butanoate	5,78E-04	Mesure expérimentale
Esters	123-86-4	butyl acetate	5,13E-02	Mesure expérimentale
Esters	141-78-6	ethyl acetate	2,91E+00	Mesure expérimentale
Esters	623-42-7	methyl butanoate	3,32E-02	Mesure expérimentale
Esters	624-24-8	methyl pentanoate	1,01E-02	Mesure expérimentale
Ethers	109-99-9	oxolane(THF)	2,26E+00	Mesure expérimentale
Ethers	1634-04-4	2-methoxy-2-methylpropane(MTBE)	2,50E-02	Mesure expérimentale
Ethers	60-29-7	ethoxyethane	1,00E-01	Mesure expérimentale
Soufrés	75-18-3	Methylsulfanylmethane(DMS)	4,00E-03	Mesure expérimentale

Sur l'ensemble des molécules utilisées, seuls douze seuils de perception n'ont pas été mesurés par Olentica mais sont issus de la littérature. Il s'agit de 8 alcanes et de 4 molécules aromatiques communément admises comme ayant des seuils de perception parmi les plus élevés. Ainsi la précision nécessaire sur la valeur du seuil de perception est moins préjudiciable.

À partir de ces molécules, plusieurs mélanges liquides sont fabriqués selon le protocole présenté ci-dessous en commençant par la molécule dont la pression de vapeur saturante est la plus basse :

- Pipetage d'un volume connu du composé pur ;
- Introduction itérative de chaque volume dans un flacon préalablement taré avec pesée du volume introduit à chaque ajout.

Une évaluation de la justesse et de la précision de cette méthode de constitution de ces mélanges liquides a été effectuée. Elle a été réalisée en constituant cinq mélanges identiques. La justesse de fabrication est évaluée à l'aide de l'erreur relative moyenne (ERM) d'après l'Équation 9.

$$ERM_i = \left(\frac{|M_{moyen} - m_{théorique}|}{m_{théorique}} \times 100 \right) \quad \text{Avec :} \quad \begin{array}{l} m_{théorique} : \text{masse théorique correspondant au volume injecté théorique de la} \\ \text{molécule } i \\ M_{moyen} : \text{moyenne des cinq masses réellement ajoutées de la molécule } i \end{array}$$

Équation 9 : Définition de l'erreur relative moyenne calculée par molécule i du mélange liquide

La précision est estimée à l'aide du coefficient de variation (CV) d'après l'Équation 10.

$$CV_i = \left(\frac{SD_{moyen}}{M_{moyen}} \times 100 \right) \quad \text{Avec :} \quad \begin{array}{l} SD_{moyen} : \text{écart-type des cinq masses réellement ajoutées de la molécule } i \\ M_{moyen} : \text{moyenne des cinq masses réellement ajoutées de la molécule } i \end{array}$$

Équation 10 : Définition du coefficient de variation calculé par molécule i du mélange liquide

La Figure 51 présente le positionnement des différentes molécules utilisées dans l'espace défini par la précision et la justesse. Les données qui ont permis d'établir cette représentation sont présentées en Annexe 21.

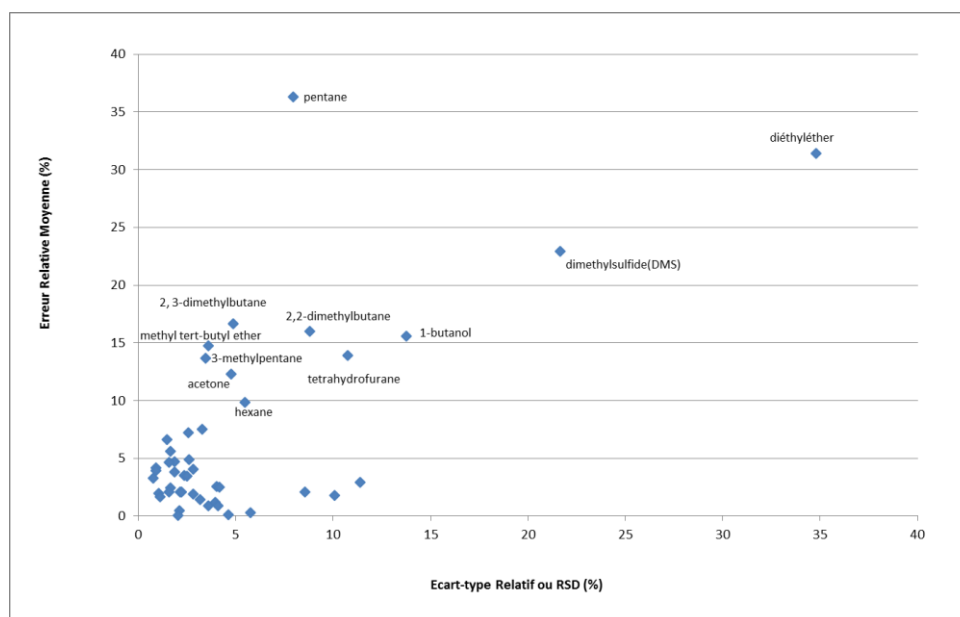


Figure 51 : Positionnement des molécules dans l'espace de justesse et précision

Les résultats de cette validation de la méthode de fabrication montrent qu'en fixant un critère de justesse et de précision supérieur à 85 %, seul six molécules se situent en dehors de cette tolérance. Pour quatre d'entre eux (pentane, 2,2-diméthylbutane, 2,3-diméthylbutane et 1-butanol), les quantités introduites dans la solution liquide sont fausses mais précises ce qui ne constitue pas un obstacle à la génération d'un mélange gazeux précis. Cet écart est lié à la pression de vapeur élevée de ces molécules. En ce qui concerne le diéthyléther (ethoxyethane) et le DMS (methylsulfanylmethane), les quantités introduites sont éloignées des quantités théoriques mais elles sont de plus très variables d'une pesée à l'autre ce qui est plus gênant pour la génération du mélange gazeux.

Cinq nouveaux mélanges différents ont été générés. Ce nombre limité s'explique par la complexité de la fabrication et le temps nécessaire à la réalisation de toutes les analyses de contrôle. Le tableau ci-

dessous détaille la composition de ces cinq mélanges liquides fabriqués pour cette étude. Les différences de nature et de quantité des molécules permettent de modifier les propriétés olfactives des mélanges.

Tableau 35 : Composition des cinq mélanges liquides synthétiques (masses pesées (mg))

N° CAS	Molécule	Famille	Mélange 1	Mélange 2	Mélange 3	Mélange 4	Mélange 5
109-66-0	pentane	Alcanes	77,6	81,5	39,6	37,1	107,7
110-54-3	hexane	Alcanes	68,1	144,2	57,6	55,5	122,6
111-65-9	octane	Alcanes	310,8	133,2	66,8	68,2	102,8
111-84-2	nonane	Alcanes	233	146,4	69,4	71,4	139,3
1120-21-4	undecane	Alcanes	83,2	142,3	70,2	74,4	148,8
124-18-5	decane	Alcanes	88,5	138,6	79,9	76,7	150
142-82-5	heptane	Alcanes	214,9	130,2	67,5	72,1	133,6
75-83-2	2,2-dimethylbutane	Alcanes	241,1	106,2	57	59,3	102,6
79-29-8	2,3-dimethylbutane	Alcanes	86,6	143	55,4	47,4	111,6
96-14-0	3-methylpentane	Alcanes	258,9	124	47,5	63,9	116,9
110-83-8	cyclohexene	Alcènes	76,8	149,3	76,9	36,9	152,1
111-66-0	oct-1-ene	Alcènes	85	137,4	7,3	0	139,8
592-41-6	hex-1-ene	Alcènes	82,8	129,3	0	0	63
592-76-7	hept-1-ene	Alcènes	86,8	132,8	64,3	33,5	133,1
104-76-7	2-ethylhexan-1-ol	Alcools	74,2	149,1	77,9	57,9	391,2
64-17-5	ethanol	Alcools	0	0	0	715,3	150,4
71-36-3	butan-1-ol	Alcools	55,3	158,6	78,1	40,7	394,5
112-31-2	decanal	Aldéhydes	26,8	156,9	7,5	0	164,1
123-05-7	2-ethylhexanal	Aldéhydes	52,9	155,9	79,8	40,3	161,3
124-13-0	octanal	Aldéhydes	0	0	0	0	161
590-86-3	3-methylbutanal	Aldéhydes	43	153,6	0	0	36,6
96-17-3	2-methylbutanal	Aldéhydes	67,3	163,2	0	0	44,2
100-41-4	ethylbenzene	Aromatiques	80,2	167	84	42,9	169,3
100-42-5	styrene	Aromatiques	83,3	170,9	86,7	859,4	176
106-42-3	1,4-xylene	Aromatiques	66,2	164,8	82,6	40,8	172,3
108-38-3	1,3-xylene	Aromatiques	83,5	165,8	84,3	44,4	169,1
108-88-3	toluene	Aromatiques	89,4	170,7	81,8	40,1	172
95-47-6	1,2-xylene	Aromatiques	81,6	168,6	88,8	42,6	175,8
95-63-6	1,2,4-trimethylbenzene	Aromatiques	72,8	167,5	88,2	45,4	169
108-10-1	4-methylpentan-2-one	Cétones	67	152	77,1	76,7	154,8
108-83-8	2,6-dimethylheptan-4-one	Cétones	67,4	155,1	76,9	78,2	155,1
110-43-0	heptan-2-one	Cétones	64,4	157,1	79,1	82,9	167,4
431-03-8	butane-2,3-dione	Cétones	0	190,9	0	0	8,9
589-38-8	hexan-3-one	Cétones	80,9	156,4	81,1	77,5	170
591-78-6	hexan-2-one	Cétones	76,9	155,9	78,2	77,6	154,1
67-64-1	propan-2-one	Cétones	268,6	137,6	55,9	72,9	142,7
78-93-3	butan-2-one	Cétones	237,2	151,7	75,3	72,3	154,2
98-86-2	1-phenylethanone	Cétones	71,1	201,3	524,3	51,9	198,9
105-37-3	ethyl propanoate	Esters	80,2	171,2	83,4	0	173,6
105-54-4	ethyl butanoate	Esters	71,1	165,5	0	0	175
123-86-4	butyl acetate	Esters	57,5	169,8	83,8	84,6	171,6
141-78-6	ethyl acetate	Esters	85,3	172,2	84,9	88	175,4
623-42-7	methyl butanoate	Esters	57,4	172,6	83,7	44,6	182,2
624-24-8	methyl pentanoate	Esters	56,4	169,7	434,7	0	176,8
109-99-9	oxolane(THF)	Ethers	212,1	161,2	76,9	79,1	168,2
1634-04-4	2-methoxy-2-methylpropane	Ethers	77,3	130,1	314,6	61,9	126,6
60-29-7	ethoxyethane	ethers	63,5	97,4	62,8	22,3	106,6
75-18-3	Methylsulfanylmethane (DMS)	Soufrés	58,8	119,7	5,2	0	155,6

Un équilibre entre les différentes familles chimiques est respecté. Le nombre de molécules par mélange varie de 37 (mélange 4) à 48 (mélange 5). La proportion de seuils de perception mesurés par mélange est présentée en Tableau 36.

Tableau 36 : Proportion de seuils de perception mesurés expérimentalement par mélange

	Mélange 1	Mélange 2	Mélange 3	Mélange 4	Mélange 5
Nombre de molécules dont le seuil de perception est mesuré par Olentica	33	34	29	25	36
Nombre total de molécules dans le mélange	45	46	41	37	48
Pourcentage de molécules dont le seuil est mesuré par Olentica	73 %	74 %	71 %	68 %	75 %

Il y a donc en moyenne 73% de seuils de perception mesurés par Olentica par mélange.

4.2.1.2 Des mélanges liquides aux mélanges gazeux synthétiques

Chaque mélange liquide synthétique a été volatilisé selon le protocole présenté en Annexe 22. La répétabilité de la génération des mélanges gazeux synthétiques ainsi que celle des analyses physico-chimiques et olfactométriques est évaluée dans un premier temps. Les mélanges synthétiques utilisés pour mettre en place la démarche automatisée sont ensuite présentés.

4.2.1.2.1 Évaluation de la répétabilité de la génération des mélanges gazeux

Une étude de la répétabilité de la génération des mélanges gazeux a été conduite selon le protocole présenté dans la figure suivante.

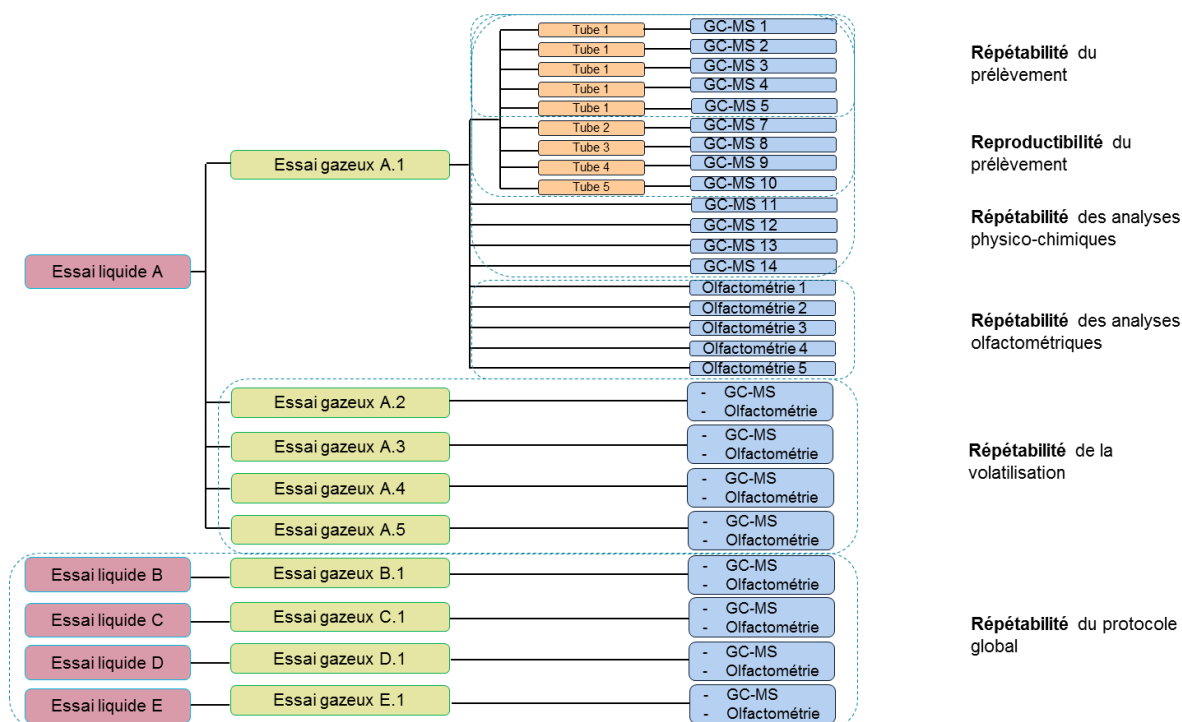


Figure 52 : Plan d'étude de la répétabilité de la fabrication d'un mélange gazeux synthétique

Un mélange liquide (essai liquide A - Annexe 23) est fabriqué. Une fraction de ce mélange est volatilisée (essai gazeux A.1).

Cet échantillon gazeux est ensuite analysé par TD-GC-MS et olfactométrie selon la norme EN 13725. Les incertitudes de mesure liées au prélèvement sur un même tube adsorbant (répétabilité du prélèvement), ou sur des tubes différents (reproductibilité), ainsi que celles liées à l'analyse physico-chimique et à l'analyse olfactométrique sont évaluées sur ce premier échantillon gazeux.

Concernant les analyses physico-chimiques, une mesure de l'incertitude est réalisée pour chacune des molécules présentes dans l'échantillon (résultats fournis en Annexe 24). Le RSD moyen représente alors la moyenne de ces RSD individuels.

À partir de cet échantillon liquide, le processus précédent est répété quatre fois. Cela aboutit à l'obtention de quatre mélanges gazeux (essais gazeux A.2 à A.5). Cette étape permet d'évaluer l'incertitude cumulée liée à la génération du mélange gazeux et à la mesure physico-chimique et olfactométrique.

Enfin, l'incertitude globale de génération d'atmosphères étalon est évaluée en répétant quatre fois la fabrication du mélange liquide (essais liquides B à E) et en effectuant une analyse physico-chimique et olfactométrique sur chaque échantillon gazeux généré à partir de ceux-ci (mélanges gazeux B.1 à E.1).

Le Tableau 37 synthétise les résultats.

Tableau 37 : Évaluation des incertitudes lors de la constitution des mélanges gazeux

Incertaince	Analyse physico-chimique (TD-GC-MS)			Analyse olfactométrique
	RSD _{minimum}	RSD _{moyen}	RSD _{maximum}	RSD
Répétabilité du prélèvement	0,5 %	4,1 %	11,5 %	
Reproductibilité du prélèvement	1,4 %	4,6 %	11,5 %	
Répétabilité des analyses physico-chimiques	0,9 %	3,7 %	7,0 %	13 %
Répétabilité de la volatilisation	5,0 %	13,0 %	21,0 %	17 %
Répétabilité du protocole global	1,5 %	8,0 %	31,0 %	18 %

Au niveau des analyses physico-chimiques, seules les concentrations de quatre molécules ont une précision inférieure à 85 %. On peut distinguer deux cas de figure :

- Le diéthyléther (ethoxyethane) et le DMS (methylsulfanylmethane) qui présentaient déjà une imprécision supérieure à 15 % lors de la constitution du mélange liquide. Il n'est donc pas étonnant de mesurer pour eux une incertitude supérieure à 15 % dans la phase gazeuse.
- Le décanal et l'acétophénone qui ont également une incertitude de mesure supérieure à 15 %. Pour ces molécules, des problèmes de séparation chromatographique et par conséquent de quantification peuvent expliquer cette incertitude plus importante.

Ces résultats permettent de valider notre protocole de génération d'un échantillon gazeux à partir d'une solution liquide.

4.2.1.2.2 Composition des mélanges de référence

La composition des cinq mélanges gazeux générés, à partir des cinq mélanges liquide du Tableau 35, pour servir de base de données pour notre étude sont présentées dans le Tableau 38.

Tableau 38 : Composition massique des 5 mélanges gazeux synthétiques (mg/m^3)

No. CAS	Molécules	Famille	Mélange 1	Mélange 2	Mélange 3	Mélange 4	Mélange 5
109-66-0	pentane	Alcanes	1,6	1,2	5,3	1,0	1,5
110-54-3	hexane	Alcanes	1,4	2,1	7,7	1,5	1,7
111-65-9	octane	Alcanes	6,5	1,9	8,9	1,9	1,4
111-84-2	nonane	Alcanes	4,9	2,1	9,3	2,0	1,9
1120-21-4	undecane	Alcanes	1,7	2,1	9,4	2,1	2,1
124-18-5	decane	Alcanes	1,8	2,0	10,7	2,1	2,1
142-82-5	heptane	Alcanes	4,5	1,9	9,0	2,0	1,9
75-83-2	2,2-dimethylbutane	Alcanes	5,0	1,5	7,6	1,7	1,4
79-29-8	2,3-dimethylbutane	Alcanes	1,8	2,1	7,4	1,3	1,6
96-14-0	3-methylpentane	Alcanes	5,4	1,8	6,3	1,8	1,6
110-83-8	cyclohexene	Alcènes	1,6	2,2	10,3	1,0	2,1
111-66-0	oct-1-ene	Alcènes	1,8	2,0	1,0	0,0	2,0
592-41-6	hex-1-ene	Alcènes	1,7	1,9	0,0	0,0	0,9
592-76-7	hept-1-ene	Alcènes	1,8	1,9	8,6	0,9	1,9
104-76-7	2-ethylhexan-1-ol	Alcools	1,5	2,2	10,4	1,6	5,5
71-36-3	butan-1-ol	Alcools	1,2	2,3	10,4	1,1	5,5
64-17-5	ethanol	Alcools	0,0	0,0	0,0	19,9	2,1
112-31-2	decanal	Aldéhydes	0,6	2,3	1,0	0,0	2,3

123-05-7	2-ethylhexanal	Aldéhydes	1,1	2,3	10,6	1,1	2,3
590-86-3	3-methylbutanal	Aldéhydes	0,9	2,2	0,0	0,0	0,5
96-17-3	2-methylbutanal	Aldéhydes	1,4	2,4	0,0	0,0	0,6
124-13-0	octanal	Aldéhydes	0,0	0,0	0,0	0,0	2,3
100-41-4	ethylbenzene	Aromatiques	1,7	2,4	11,2	1,2	2,4
100-42-5	styrene	Aromatiques	1,7	2,5	11,6	24,0	2,5
106-42-3	1,4-xylene	Aromatiques	1,4	2,4	11,0	1,1	2,4
108-38-3	1,3-xylene	Aromatiques	1,7	2,4	11,2	1,2	2,4
108-88-3	toluene	Aromatiques	1,9	2,5	10,9	1,1	2,4
95-47-6	1,2-xylene	Aromatiques	1,7	2,4	11,8	1,2	2,5
95-63-6	1,2,4-trimethylbenzene	Aromatiques	1,5	2,4	11,8	1,3	2,4
108-10-1	4-methylpentan-2-one	Cétones	1,4	2,2	10,3	2,1	2,2
108-83-8	2,6-dimethylheptan-4-one	Cétones	1,4	2,2	10,3	2,2	2,2
110-43-0	heptan-2-one	Cétones	1,3	2,3	10,5	2,3	2,3
589-38-8	hexan-3-one	Cétones	1,7	2,3	10,8	2,2	2,4
591-78-6	hexan-2-one	Cétones	1,6	2,3	10,4	2,2	2,2
67-64-1	propan-2-one	Cétones	5,6	2,0	7,5	2,0	2,0
78-93-3	butan-2-one	Cétones	4,9	2,2	10,0	2,0	2,2
98-86-2	1-phenylethanone	Cétones	1,5	2,9	69,9	1,4	2,8
431-03-8	butane-2,3-dione (diacetyl)	Cétones	0,0	2,8	0,0	0,0	0,1
105-37-3	ethyl propanoate	Esters	1,7	2,5	11,1	0,0	2,4
105-54-4	ethyl butanoate	Esters	1,5	2,4	0,0	0,0	2,4
123-86-4	butyl acetate	Esters	1,2	2,5	11,2	2,4	2,4
141-78-6	ethyl acetate	Esters	1,8	2,5	11,3	2,5	2,5
623-42-7	methyl butanoate	Esters	1,2	2,5	11,2	1,2	2,5
624-24-8	methyl pentanoate	Esters	1,2	2,5	58,0	0,0	2,5
60-29-7	ethoxyethane	Ethers	1,3	1,4	8,4	0,6	1,5
109-99-9	oxolane(THF)	Ethers	4,4	2,3	10,3	2,2	2,4
1634-04-4	2-methoxy-2-methylpropane(MTBE)	Ethers	1,6	1,9	42,0	1,7	1,8
75-18-3	Methylsulfanylmethane(DMS)	Soufrés	1,2	1,7	0,7	0,0	2,2

À l'issue de cette étude, nous disposons donc de cinq mélanges gazeux que nous pouvons utiliser comme base de données d'entrée pour mettre au point le protocole automatisé d'identification des molécules responsables de l'odeur.

4.2.2 Approche « expert » : Détermination des molécules responsables de l'odeur des mélanges

Dans un premier temps, à partir des mélanges précédents, l'expert calcule les activités d'odeur de chaque mélange, comme présenté dans le Tableau 39.

Tableau 39 : Activités d'odeur des molécules pour les cinq mélanges gazeux synthétiques (en grisé, les activités d'odeur inférieure à 1)

Molécules	Mélange 1	Mélange 2	Mélange 3	Mélange 4	Mélange 5
2,3-dimethylbutane	1	1	5	1	1
2,2-dimethylbutane	0	0	0	0	0
3-methylpentane	0	0	0	0	0
decane	0	0	0	0	0
heptane	0	0	0	0	0
hexane	0	0	1	0	0
nonane	0	0	1	0	0
octane	1	0	1	0	0
pentane	0	0	1	0	0
undecane	0	0	1	0	0
hept-1-ene	13	14	60	7	13
hex-1-ene	3300	3587	0	0	1686
oct-1-ene	247	278	136	0	272
cyclohexene	7	9	44	4	9
butan-1-ol	22	44	200	22	106
2-ethylhexan-1-ol	4	6	27	4	14
ethanol	0	0	0	213	22
2-ethylhexanal	14	29	138	15	29
2-methylbutanal	836	1411	0	0	369
3-methylbutanal	4070	10125	0	0	2327
decanal	199	810	356	0	817
octanal	0	0	0	0	4503
1,2,4-trimethylbenzene	2	4	18	2	4
ethylbenzene	5	8	35	4	7
1,3-xylene	3	5	22	2	5
1,2-xylene	1	2	8	1	2

1,4-xylene	2	4	19	2	4
styrene	29	42	195	405	42
toluene	2	2	9	1	2
butan-2-one	2	1	4	1	1
heptan-2-one	17	29	135	30	30
hexan-2-one	3	4	17	4	4
butane-2,3-dione	0	23564	0	0	1060
2,6-dimethylheptan-4-one	3	6	26	5	5
hexan-3-one	15	21	99	20	22
4-methylpentan-2-one	2	3	14	3	3
propan-2-one	1	0	1	0	0
1-phenylethanone	114	225	5399	112	215
ethyl acetate	1	1	4	1	1
butyl acetate	23	48	218	46	47
ethyl propanoate	76	113	507	0	111
ethyl butanoate	2559	4148	0	0	4231
methyl pentanoate	117	244	5758	0	246
methyl butanoate	36	75	336	37	77
ethoxyethane	13	14	84	6	15
2-methoxy-2-methylpropane (MTBE)	64	76	1680	69	71
oxolane (THF)	2	1	5	1	1
Methylsulfanylmethane (DMS)	306	434	173	0	544

La majorité des alcanes ont une activité d'odeur inférieure à 1 ce qui confirme leur très faible poids dans la concentration d'odeur des mélanges.

La littérature stipule le plus souvent que la somme des activités d'odeur ainsi calculée est à comparer à la concentration de l'odeur mesurée par olfactométrie [(CAPELLI et al. 2013) ; (WU et al. 2016) ; (Wu et al. 2017)]. Cependant, certains auteurs s'interrogent sur l'utilisation de la somme des activités d'odeur ou de l'activité d'odeur maximale correspondant à la molécule ayant le rapport concentration dans l'air sur seuil de perception le plus important [(BLAZY et al. 2015) ; (RINCON et al. 2019)].

Le Tableau 40 présente les concentrations d'odeur mesurées, la somme des activités d'odeur (AO_{som}) et l'activité d'odeur maximale de chacun des mélanges (AO_{max}).

Tableau 40 : Comparaison des concentrations d'odeur mesurées avec la somme des activités d'odeur (AO_{som}) et l'activité d'odeur maximale (AO_{max})

Échantillons	Concentration d'odeur (CO) expérimentale (UO_E/m^3)	Somme des activités d'odeur (AO_{som})	Erreur relative (%)	Max des activités d'odeur (AO_{max})	Erreur relative (%)
Mélange 1	13798	12114	12	4070	71
Mélange 2	31175	45389	46	23564	24
Mélange 3	9950	15738	58	5758	42
Mélange 4	788	1018	29	405	49
Mélange 5	17653	16918	4	4503	74

Pour la majorité de nos mélanges la somme des activités d'odeur apparaît plus proche de la concentration d'odeur que l' AO_{max} .

L'erreur relative entre les deux valeurs varie de 4% à 58% avec une moyenne de 30% et un écart-type de 20% alors que l'erreur relative pour l'activité d'odeur maximum varie entre 24% et 74% avec une moyenne de 52% et un écart-type de 19%.

Cette bonne concordance nous permet donc de considérer que si la somme des activités d'odeur est de l'ordre de grandeur de la concentration d'odeur mesurée, alors les molécules identifiées doivent être celles qui expliquent cette odeur. À l'inverse, si cette règle n'est pas vérifiée, alors l'identification des

molécules est sans doute incomplète et la tentative d'explication de la concentration d'odeur est à prendre avec la plus grande prudence.

La proximité constatée entre la somme des activités d'odeur et la concentration d'odeur mesurée par olfactométrie va nous permettre de créer de nouveaux mélanges « virtuels ». Ils seront créés en attribuant arbitrairement à chaque mélange une concentration d'odeur théorique respectant une erreur relative moyenne de 30% entre la somme des activités d'odeur des molécules constituant le mélange et la valeur de concentration d'odeur théorique attribuée.

Ces mélanges seront établis de telle sorte que la nature des molécules expliquant la concentration d'odeur soit différente de celle des mélanges réels. Cela permettra de disposer d'une plus grande quantité de données, d'une plus grande variabilité de profils olfactifs et de tester plus largement la robustesse de la démarche.

Sur la base des résultats précédents, six mélanges virtuels ont donc été imaginés. Leurs activités d'odeur sont présentées en Tableau 41.

Tableau 41 : Composition et activités d'odeur des mélanges virtuels

Molécules	Mélange 6	Mélange 7	Mélange 8	Mélange 9	Mélange 10	Mélange 11
2-methylbutanal	7000	45	12	81	46	4873
Methylsulfanylmethane(DMS)	3500	36	15000	9	78	10
2-ethylhexan-1-ol	1550	3000	66	4128	34	34
butyl acetate	1200	0	12	3456	35	1006
2-ethylhexanal	900	65	1	1543	0	27
ethyl acetate	890	5679	25	246	2369	1600
butan-1-ol	790	0	1	15	13	1
oct-1-ene	764	1264	0	0	0	0
decanal	269	13	0	65	1	0
hexan-3-one	65	215	6364	67	51	465
hex-1-ene	42	0	1	10	100	35
3-methylbutanal	30	0	45	3	13	218
ethyl butanoate	12	32	16	986	1354	3946
4-methylpentan-2-one	12	0	5	1	96	2
hexan-2-one	11	7834	35	563	19	1
octanal	10	1	981	35	13	4
1,2,4-trimethylbenzene	9	351	32	0	16	1
methyl pentanoate	7	0	6	19	1	731
ethylbenzene	6	1	17	13	95	0
decane	6	0	684	35	158	0
heptan-2-one	3	1	123	194	0	3
hept-1-ene	3	347	0	654	21	1
1,3-xylene	3	325	16	1000	77	3
propan-2-one	3	0	21	64	1	22
octane	3	1	0	0	15	0
2-methoxy-2-methylpropane (MTBE)	2	0	2	30	0	1546
ethoxyethane	2	0	1	2	0	2
cyclohexene	2	120	5	0	84	1
2,6-dimethylheptan-4-one	2	0	0	646	983	37
methyl butanoate	1	1594	61	264	3	2640
ethanol	1	0	49	0	3	450
1,4-xylene	1	1	0	0	1	1
butan-2-one	1	2300	0	0	764	864
nonane	1	0	3	1	4	1
3-methylpentane	1	2	33	0	1	0
butane-2,3-dione	0	377	0	2497	0	13
1-phenylethanone	0	0	3	1	55	361
ethyl propanoate	0	10	791	33	0	1340
styrene	0	0	1	0	0	213
toluene	0	0	0	0	0	1
1,2-xylene	0	0	5	10	1	3
2,3-dimethylbutane	0	10	776	510	67	1
oxolane (THF)	0	0	0	0	1	0
pentane	0	0	0	1	0	0
undecane	0	1	0	0	0	0
hexane	0	2	3	0	0	1
heptane	0	0	1	0	0	0
2,2-dimethylbutane	0	1	3000	3	0	0

La somme des activités d'odeur, la concentration d'odeur attribuée et le pourcentage de variation sont présentés pour chacun des mélanges virtuels en Tableau 42.

Tableau 42 : Caractéristiques des mélanges virtuels

Echantillons	Concentration d'odeur attribuée (CO) (UO_E/m^3)	(AO_{Som})	Erreur relative (%)
Mélange 6	20000	17102	14
Mélange 7	24980	23628	5
Mélange 8	35460	28197	20
Mélange 9	21646	17185	21
Mélange 10	7294	6573	10
Mélange 11	17694	20458	16

L'écart entre la concentration d'odeur et la somme des activités d'odeur respecte la variation observée pour les mélanges réels.

À partir de l'ensemble de ces résultats, l'expert identifie les molécules responsables de la concentration d'odeur des cinq mélanges synthétiques et des six mélanges virtuels.

Pour son analyse, il considère les données des différentes matrices gazeuses individuellement, c'est-à-dire qu'il va comparer les activités d'odeur des molécules avec la concentration d'odeur mesurée ou attribuée pour le mélange. Il calcule les activités d'odeur sur la base des concentrations physico-chimiques et des valeurs de seuils de perception dont il dispose.

Les molécules considérées par l'expert comme responsables de l'odeur des onze mélanges sont présentées dans le Tableau 43.

Tableau 43 : Molécules responsables de l'odeur de chaque mélange selon l'expert

Mélanges	Molécules par ordre d'importance dans la concentration de l'odeur						
Mélanges synthétiques	1	3-methylbutanal	hex-1-ene	ethyl butanoate	2-methylbutanal		
	2	butane-2,3-dione	3-methylbutanal	ethyl butanoate	hex-1-ene		
	3	methyl pentanoate	1-phenylethanone	2-methoxy-2-methylpropane (MTBE)			
	4	styrene	ethanol	1-phenylethanone	2-methoxy-2-methylpropane (MTBE)		
	5	octanal	ethyl butanoate	3-methylbutanal	hex-1-ene	butane-2,3-dione	
Mélanges virtuels	6	2-methylbutanal	methylsulfanylmethane (DMS)	2-ethylhexan-1-ol	butyl acetate	2-ethylhexanal	ethyl acetate
	7	2-hexanone	ethyl acetate	2-ethylhexan-1-ol	2-butanone	methyl butanoate	1-octene
	8	methylsulfanylmethane (DMS)	3-hexanone	2,2-dimethylbutane			
	9	2-ethylhexan-1-ol	ethyl acetate	butane-2,3-dione	2-ethylhexanal	1,3-xylene	ethyl butanoate
	10	ethyl acetate	ethyl butanoate	2,6-dimethylheptan-4-one	2-butanone		
	11	2-methylbutanal	ethyl butanoate	Methylbutyrate	ethyl acetate	2-methoxy-2-methylpropane (MTBE)	ethyl propanoate

Ces résultats permettent de constater la grande diversité des molécules responsables de l'odeur des onze mélanges. Ces résultats seront par la suite utilisés comme base de comparaison pour valider la démarche automatisée.

4.2.3 Démarche automatisée d'identification des molécules responsables de l'odeur

L'objectif de la démarche automatisée est double :

- Calculer automatiquement les activités d'odeur de l'échantillon ciblé et donc retrouver rapidement les résultats de l'expert ;
- Identifier, à l'aide d'une régression, les molécules expliquant le plus une variation de la concentration d'odeur. En effet, cette approche permet de vérifier si les molécules identifiées comme importantes dans l'explication de l'odeur par l'expert sont également celles qui impactent le plus une variation de la concentration d'odeur.

Si l'atteinte du premier objectif ne nécessite qu'une succession de manipulations de programmation, celle du second nécessite d'identifier la régression statistique qui permettra le mieux de répondre au problème posé.

Les données nécessaires pour réaliser cette régression sont identiques aux données utilisées par l'expert :

- La nature des molécules présentes dans l'échantillon analysé et leur concentration exacte (en mg/m^3) ;
- Les valeurs de seuils de perception (en mg/m^3) des molécules présentes ;
- La concentration d'odeur de l'échantillon analysé

À la différence de l'expert, il est nécessaire de disposer de ces données pour un ensemble d'échantillons minimal pour envisager d'appliquer la régression. Cela permet d'observer la covariance de l'activité d'odeur des molécules (X) avec la variable réponse (la concentration d'odeur : Y), autrement dit, leur sens d'évolution les unes par rapport aux autres qui n'est pas observable dans l'approche experte.

Le format des données utilisées pour la régression est présenté en Figure 53.

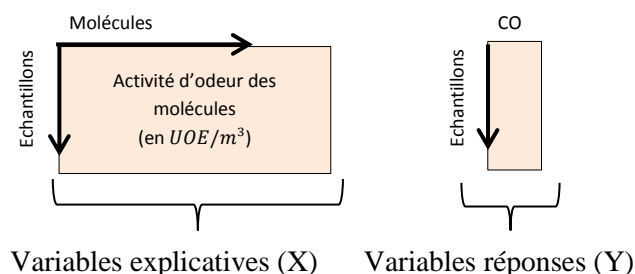


Figure 53 : Format des données utilisées dans l'approche statistique

La suite de ce chapitre est consacrée au choix du modèle de régression en fonction des critères imposés par la nature des données d'entrées. La première étape est donc la définition de ces contraintes.

4.2.3.1 Contraintes liées aux données existantes

Un premier point clé de l'approche est le choix des études à croiser avec l'échantillon analysé, pris comme référence.

L'obtention d'un modèle stable implique, d'un point de vue statistique, de croiser des données présentant le maximum de variables explicatives communes.

Dans notre cas, il est nécessaire de rechercher les échantillons présentant le plus grand nombre de molécules communes avec l'échantillon analysé de référence. Ce premier critère n'est toutefois pas suffisant pour s'assurer que le modèle représente fidèlement le phénomène étudié. En effet, même si les molécules présentes sont identiques dans tous les échantillons, celles qui expliquent la concentration d'odeur peuvent être différentes d'un échantillon à l'autre. Il est donc nécessaire d'introduire un second critère de similarité. Ce critère est calculé à partir du nombre de molécules ayant une activité d'odeur représentant plus de 1% de la somme des activités d'odeur de l'échantillon considéré.

Un critère de similarité global est obtenu en utilisant le produit de ces deux pourcentages. La valeur seuil de ce critère qui permettra à la régression de fournir le résultat le plus cohérent avec la réalité est choisie en testant différents niveaux de similarité pour les cinq mélanges synthétiques présentés au paragraphe 4.2.1.2.

Un deuxième point clé, qu'il est très important de souligner ici est, que dans la pratique courante, le nombre de variables (le nombre de molécules) est quasi systématiquement très supérieur au nombre d'échantillons qu'il est possible d'utiliser dans une même étude. En effet, il ne faut pas perdre de vue que le nombre d'échantillons utilisables dans une étude donnée est contraint par la nécessité de respecter le critère de similarité présenté ci-dessus.

Enfin, un troisième point clé est la contrainte de mettre en relation une quantité importante de données physicochimiques précises avec des données sensorielles beaucoup plus imprécises et en faible nombre. Le choix de la méthode de régression ci-dessous intègre donc ces différents niveaux de contrainte.

4.2.3.2 Choix du type de régression statistique

L'une des régressions adaptées aux contraintes liées à nos données est la régression des moindres carrés partiels plus communément appelée régression PLS pour « Partial Least Square regression » (WOLD, SJOSTROM, et ERIKSSON 2001). Cette méthode a fait ses preuves dans les domaines de la chimie (TENENHAUS 1998c), de l'agroalimentaire ou de la cosmétique [(TUFFERY 2010) ; (GILBERT 2012) ; (PICARD 2013)].

Elle présente de multiples avantages par rapport à une régression linéaire classique [(DECREMER 2006) ; (BINARD 2013)]:

- C'est une méthode robuste qui permet de gérer les cas pour lesquels le nombre d'échantillons est inférieur au nombre de variables [(CRAMER III 1993) ; (TENENHAUS 1998c) ; (TENENHAUS 1998a)].
- Elle est stable même lorsque deux variables sont fortement corrélées (TENENHAUS, GAUCHI, et MENARDO 1995). En effet, elle n'élimine pas obligatoirement une variable qui est la combinaison linéaire d'une autre et qui serait importante. Elle pallie ainsi la multicollinéarité des données (CHONG et JUN 2005).
- Elle permet de ne pas prioriser la variance des variables explicatives ou la variable réponse pour obtenir un modèle le plus complet possible. Elle est construite de manière à trouver un compromis entre l'explication de l'une ou l'autre.
- Elle permet de calculer un indicateur d'importance pour chaque variable appelé la VIP ou « Variable Importance Projection (TENENHAUS 1998b) qui est une aide à l'interprétation des résultats : plus une variable a une forte valeur de VIP, plus cette variable est importante. Les variables ayant une valeur de VIP supérieure à 1 sont les plus importantes dans la construction de la réponse, celles ayant une VIP inférieure à 0,8 ont une importance moindre dans l'explication de la réponse [(TENENHAUS 1998b) ; (BALAKRISHNAN et al. 2012)].

Une technique concurrente de la PLS est l'approche « Random Forests » (BREIMAN 2001) puisqu'il a été montré dans une étude récente que les deux techniques présentent les mêmes performances dans la détection des variables importantes (LEE et al. 2018), malgré un sur apprentissage de PLS par rapport à « Random Forests » (FLIGHT 2015). Cependant, la technique « Random Forests » n'est pas adaptée dans le cas d'un faible nombre d'échantillons puisqu'elle est basée sur une démarche de bootstrap (BREIMAN 2001) ce qui n'est pas adapté à notre contexte.

La méthode PLS a donc été retenue dans le cadre de cette étude.

Pour construire le modèle PLS le plus pertinent, le nombre optimal de composantes doit être déterminé.

Pour cela, plusieurs indicateurs sont utilisés :

- Tout d'abord, il faut observer le pourcentage de variance de réponse Y restitué par chaque composante « t » de PLS. Le minimum de composantes permettant d'expliquer le maximum de variance de Y doit être choisi.
- Ensuite, il faut calculer l'erreur quadratique moyenne de prédiction plus communément appelée RMSEP pour « Root Mean Squared Error of Prediction ». Cette grandeur est calculée à partir de la somme des erreurs de prédiction ou PRESS pour « PRedictive Error Sum of Squares » (CHONG et JUN 2005). C'est une valeur obtenue par validation croisée [(TENENHAUS 1998c) ; (CHAVENT et PATOUILLE 2003) ; (DECREMER 2006) ; (LOHNINGER 2012)]
- Un autre indicateur est calculé à l'aide d'une validation croisée, il s'agit de la somme des erreurs au carré ou RSS pour « Residual Sum of Squares ».

- Les valeurs de PRESS et RSS permettent de calculer un quatrième indicateur nommé Q^2 aidant à choisir le nombre de composantes [(TENENHAUS 1998c) ; (DECREMER 2006)]. Il faut un minimum de 4 échantillons pour que la validation croisée ait un sens.

Le nombre de composantes à conserver correspond à la première valeur de RMSEP minimale [(LOHNINGER 2012) ; (GONZALEZ et TENENHAUS 2000)]. De plus chaque composante ayant un Q^2 supérieur ou égale à 0,0975 (TENENHAUS 1998c) est considérée comme explicative.

4.2.3.3 La démarche d'analyse automatisée d'un échantillon

Une fois l'échantillon à analyser sélectionné, il devient l'échantillon de référence et la démarche à suivre est divisée en 6 étapes.

1. La première consiste à calculer l'activité d'odeur de toutes les molécules présentes dans l'échantillon de référence afin d'identifier rapidement les molécules dont l'activité d'odeur représente plus de 5% de l'activité d'odeur globale de ce mélange et de retrouver l'identification de l'expert.
2. La deuxième étape est la constitution de la base de données pour analyser l'échantillon de référence à l'aide de l'approche statistique. Cela nécessite d'inventorier tous les échantillons disponibles (ici les 5 mélanges synthétiques et les 6 mélanges virtuels) et d'appliquer le critère de similarité adéquat tout en conservant un minimum de 4 échantillons dans cette base de données. Cette valeur permet de lancer la régression PLS. C'est également un nombre d'échantillons souvent disponible dans la pratique pour une même étude. Dans ce dernier cas, les résultats ne seront pas stables, mais il restera possible d'analyser les covariances du jeu de données et de calculer la valeur VIP pour chaque variable. Évidemment, plus le nombre d'échantillons sera important plus la stabilité du modèle sera améliorée sous réserve que les échantillons étudiés présentent un degré de similarité suffisant.
3. La troisième étape est le calcul de l'activité d'odeur de toutes les molécules présentes dans les échantillons sélectionnés pour l'étude. Cette étape met en forme les données pour affiner le résultat de l'expert par analyse PLS.
4. La quatrième étape consiste au prétraitement des données. Le centrage de l'ensemble des données est obligatoire pour que la PLS conserve ses propriétés mathématiques (DECREMER 2006). Il est donc choisi de centrer les données X et Y. La standardisation globale des données n'est pas nécessaire car les variables ont toutes la même unité. De plus elle n'est pas utile car cela annihile la pondération par les seuils de perception et implique de traiter des corrélations entre les variables explicatives et la réponse plutôt que des covariances, ce qui n'est pas recherché.
5. La cinquième étape consiste à établir une régression PLS entre les activités d'odeur et la concentration d'odeur et à choisir le nombre optimal de composantes comme présenté en partie 4.2.3.2.
6. La sixième étape est le calcul de la VIP de chaque variable pour ne conserver que celles possédant une VIP supérieure à 0,8 (cf partie 4.2.3.2). De cette façon il est possible d'identifier les molécules les plus importantes dans l'explication de la variation de la concentration d'odeur.

4.3 Résultats

Après la présentation de la méthodologie, les paragraphes suivants exposent, pour le premier, les résultats de l'application de la démarche à l'un des mélanges synthétiques, et pour le second le choix du critère de similarité conduisant aux résultats PLS les plus réalistes.

4.3.1 Mise en pratique et paramétrage de la démarche automatisée

Le mélange synthétique choisi pour exposer la méthodologie est le mélange 5.

La première étape de la démarche automatisée consiste à calculer les activités d'odeur de chacune des molécules puis le pourcentage qu'elles représentent par rapport à la somme des activités d'odeur. Ces résultats sont représentés en Figure 54.

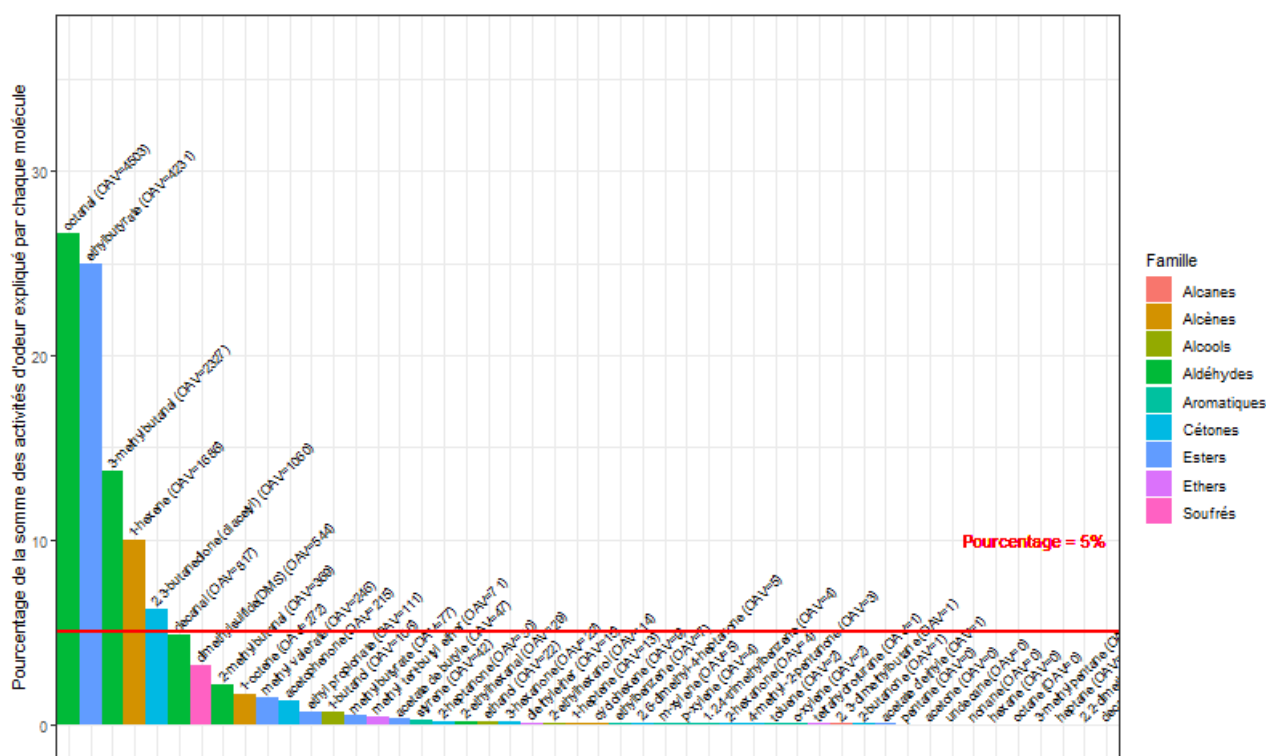


Figure 54 : Proportion de l'activité d'odeur de chaque molécule du mélange 5 par rapport à la somme des activités d'odeur de ce mélange

Dans cette application, retenir les molécules représentant plus de 5% de la concentration d'odeur globale conduit à la même identification que celle de l'expert.

La seconde étape vise à constituer la base de données nécessaire à l'analyse statistique. Pour cela, le critère de similarité du mélange 5 avec les autres mélanges est calculé et présenté en Tableau 44.

Tableau 44 : Critères de similarité du mélange 5 avec les autres mélanges

Echantillons	% de molécules communes (P1)	% de molécules dont AO>1% communes (P2)	Critère de similarité (P1 x P2)
Mélange 5	100	100	10000
Mélange 1	94	64	6016
Mélange 2	96	55	5280
Mélange 11	100	45	4500
Mélange 3	85	36	3060
Mélange 6	73	36	2628
Mélange 10	71	27	1917
Mélange 8	75	18	1350
Mélange 9	71	18	1278
Mélange 7	58	18	1044
Mélange 4	77	9	693

Retenir quatre échantillons revient à choisir un critère de similarité de 4000 et à poursuivre ainsi l'analyse avec les mélanges 5, 1, 2 et 11. L'activité d'odeur des molécules de chacun de ces échantillons est alors calculée automatiquement pour appliquer par la suite la PLS. Il s'agit de la troisième étape.

Le prétraitement décrit en partie 4.2.3.3 est appliqué puis une régression est établie entre les activités d'odeur des molécules et la concentration d'odeur des échantillons (Tableau 45). Il s'agit respectivement des variables explicatives nommées « X » et de la variable réponse nommée « Y » comme défini au paragraphe 4.2.3. Ce traitement correspond aux étapes 4 et 5 de la démarche d'analyse.

Echantillons	CO (UO_F/m^3)
Mélange 1	13798
Mélange 2	31175
Mélange 5	17653
Mélange 11	17694

Tableau 45 : Concentration d'odeur expérimentale des échantillons de l'analyse

Il faut à présent choisir le nombre de composantes du modèle. Pour cela, l'évolution de l'erreur quadratique moyenne de prédiction (RMSEP) en fonction du nombre de composantes PLS est présentée en Figure 55.

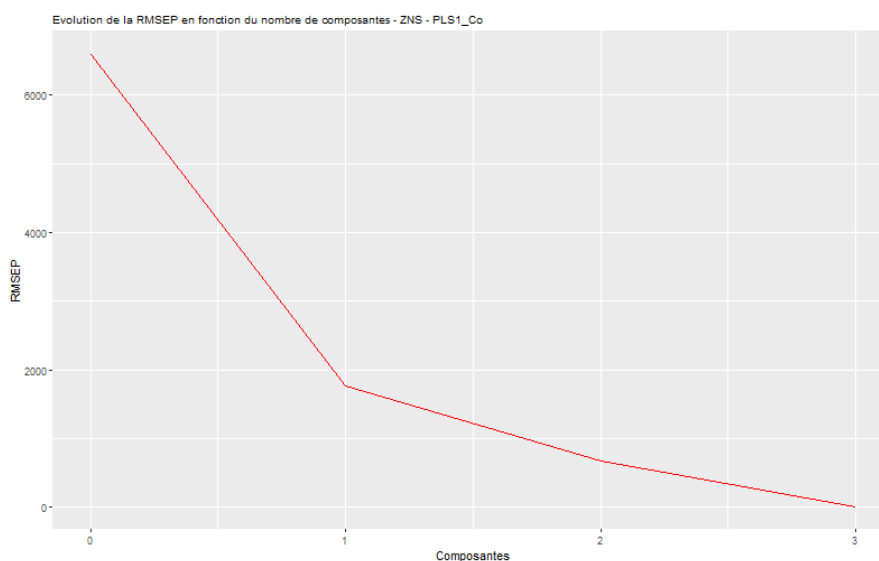


Figure 55 : Evolution de la RMSEP en fonction du nombre de composantes PLS utilisé

La Figure 55 encouragerait à choisir deux composantes. Ce résultat est confirmé par l'observation du pourcentage de variance de Y expliquée par chaque composante orthogonale « t » du modèle PLS (Tableau 46).

Tableau 46 : Pourcentage de variance de Y expliqué par les composantes orthogonales « t » de PLS

	Composante t1	Composante t2	Composante t3
Pourcentage de variance de Y expliqué par les composantes t	92,89	6,07	1,04
Cumul des variances d'explication de Y	92,89	98,96	100

Ces observations conduisent à sélectionner deux composantes, qui, cumulées, expliquent 98,96% de la variance de Y, pour construire le modèle PLS.

La projection de l'importance des variables (VIP) de chaque molécule est ensuite calculée et ces valeurs sont présentées dans la Figure 56 pour les molécules ayant une activité d'odeur représentant plus de 1% de la concentration d'odeur globale du mélange.

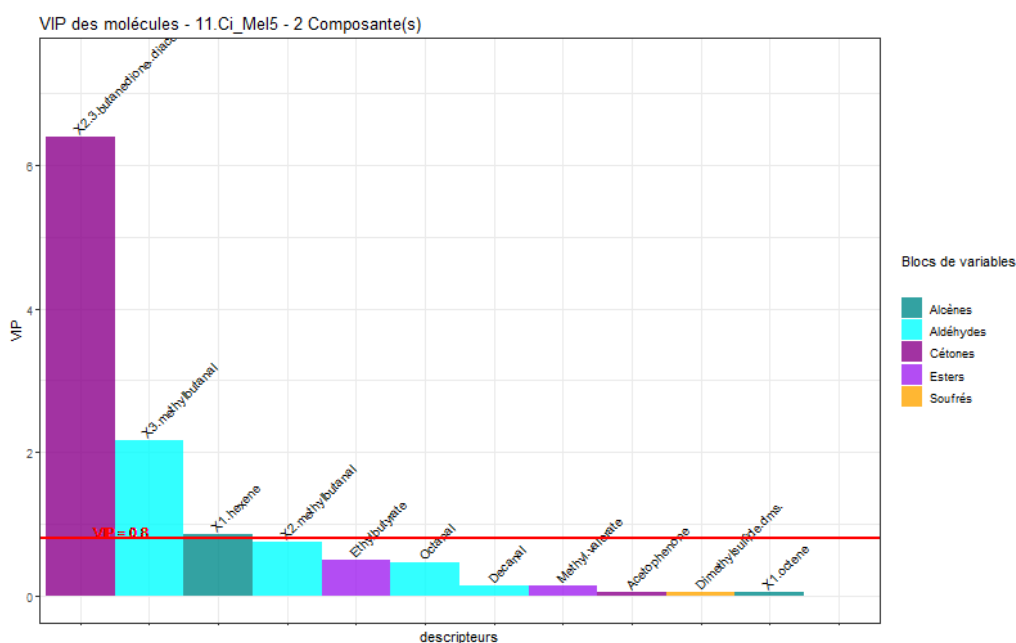


Figure 56 : VIP des molécules de l'étude menée sur le mélange 5 en référence et un critère de similarité de 4000

Parmi les 48 molécules contenues dans le mélange 5, trois d'entre elles présentent une VIP supérieure à 0,8. Parmi ces trois molécules, on constate la présence de trois des cinq molécules identifiées par l'expert mais avec un ordre d'importance différent. Le butane-3,2-dione est indiquée comme la plus importante par la méthode statistique alors qu'elle correspond à la molécule classée cinquième dans le choix de l'expert dans son analyse. Parallèlement, la molécule expliquant la part la plus importante de l'odeur du mélange étudié par l'expert n'est pas retenue par l'approche statistique.

Ce résultat montre donc une concordance partielle entre l'approche statistique et l'approche de l'expert. Il montre également que l'approche statistique fait ressortir une molécule n'expliquant pas forcément la part la plus importante de l'odeur du mélange étudié, mais dont une faible variation de la concentration serait susceptible de faire varier la concentration d'odeur de manière importante.

Il est important de souligner qu'une étude conduite, comme ici, avec le minimum nécessaire de quatre échantillons se traduit par un modèle obligatoirement instable. Le renforcement du modèle nécessite d'employer un plus grand nombre d'échantillons, donc nécessairement d'abaisser le critère de similarité. C'est pourquoi le paragraphe suivant consiste à évaluer la valeur du critère de similarité permettant d'obtenir un résultat conforme à la réalité.

4.3.2 Impact de la dégradation du critère de similarité

Pour cette partie de l'étude, la base de données des mélanges est constituée de onze mélanges : les cinq mélanges gazeux expérimentaux et les six mélanges virtuels. L'évaluation de l'impact de la dégradation du critère de similarité est réalisée pour les cinq mélanges gazeux expérimentaux en considérant cinq valeurs de critère de similarité différentes (0, 500, 1200, 2500, 4000) (résultats PLS en Annexe 25).

Une attention particulière est portée tout au long de cette étude à privilégier un modèle de régression permettant d'expliquer le plus de variation avec le moins de composantes possibles, selon le principe de parcimonie. D'après les résultats présentés en Annexe 25, on constate que le nombre de composantes nécessaire à l'explication de la variation de la variable réponse diminue avec l'augmentation du critère de similarité.

Les molécules identifiées par l'expert comme majoritairement responsables de la concentration d'odeur sont ensuite comparées à celles identifiées par l'approche statistique, et ce pour les 5 valeurs de critère de similarité testées (Tableau 47).

Tableau 47 : Comparaison de l'approche de l'expert et de l'approche statistique pour différentes valeurs de critère de similarité

Mélanges	Mélange 1					Mélange 2					Mélange 3					Mélange 4					Mélange 5				
Critère de similarité	0	500	1200	2500	4000	0	500	1200	2500	4000	0	500	1200	2500	4000	0	500	1200	2500	4000	0	500	1200	2500	4000
Nombre de mélanges	11	10	7	7	5	11	9	8	4	4	11	11	8	5	4	11	8	5	4	3	11	11	9	6	4
Nombre de molécules retenues par l'expert	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
Nombre de molécules retenues par la PLS	4	2	4	4	3	4	3	3	3	3	4	4	4	3	4	3	3	3	4	4	8	8	5	5	3
Nombre de composantes	4	3	3	3	3	4	2	3	2	2	4	4	3	2	2	4	3	2	2	2	4	4	4	2	2
Nombre de molécules en commun ente les deux approches	3	1	3	3	3	3	2	2	3	3	2	2	2	2	3	1	1	1	1	2	4	4	2	2	3
Pourcentage de molécules en commun ente les deux approches par rapport au résultat de l'expert	75	25	75	75	75	75	50	50	75	75	67	67	67	67	100	25	25	25	25	50	80	80	40	40	60

Pour les mélanges 3 et 4, on constate qu'augmenter la valeur du critère de similarité permet de faire converger l'identification obtenue par l'approche statistique avec celle de l'expert. La meilleure identification est obtenue pour une valeur de 4000. Cependant, pour le mélange 4, cette valeur de seuil restreint le nombre de mélanges utilisables à trois.

La même constatation peut être faite pour les mélanges 1, 2 et 5. Cependant, on peut noter dans ce cas qu'une valeur de critère de similarité fixée à 0 conduit à des pourcentages de molécules communes entre les deux approches aussi élevés que ceux obtenus avec des critères de similarité élevés.

Il convient toutefois de souligner qu'abaisser la valeur du critère de similarité implique d'inclure dans la construction du modèle des molécules appartenant à un plus grand nombre d'échantillons. Les modèles ainsi construits sont donc moins parcimonieux, puisqu'il faut augmenter le nombre de composantes pour expliquer la même variable.

Les résultats obtenus avec une valeur de critère de similarité de 0 sont donc entachés du bruit engendré par les échantillons utilisés.

En résumé, le choix de la valeur du critère de similarité se fait en minimisant le nombre de composantes (pour respecter le principe de parcimonie) et en maximisant le nombre de molécules communes entre les deux approches.

Dans notre cas, ce compromis nous conduit à retenir la valeur de 2500.

Afin d'observer l'homogénéité des études avec le critère de similarité de 2500, on compare les concentrations d'odeur des différentes études pour cette valeur en Tableau 48.

Tableau 48 : Concentrations d'odeur des différentes études pour un critère de similarité de 2500

Etude du mélange 1		Etude du mélange 2		Etude du mélange 3		Etude du mélange 4		Etude du mélange 5	
Mélange 1	13798	Mélange 1	13798	Mélange 3	9950	Mélange 3	9950	Mélange 1	13798
Mélange 2	31175	Mélange 2	31175	Mélange 4	788	Mélange 4	788	Mélange 2	31175
Mélange 3	9950	Mélange 5	17653	Mélange 5	17653	Mélange 11	17694	Mélange 3	9950
Mélange 5	17653	Mélange 11	17694	Mélange 11	17694	Mélange 9	21646	Mélange 5	17653
Mélange 10	7294			Mélange 6	20000			Mélange 11	17694
Mélange 11	17694							Mélange 6	20000
Mélange 6	20000								

Ce tableau permet de constater qu'il est nécessaire d'inclure dans les études construites pour analyser les mélanges 3 et 4 des échantillons ayant une concentration d'odeur plus hétérogène que pour le cas des trois autres mélanges.

Nous avons cherché à évaluer la variabilité de la concentration d'odeur pour chacune de ces études. Pour cela nous nous appuyons sur la variabilité de concentration d'odeur acceptable lors d'une mesure de la concentration d'odeur d'un échantillon selon la norme EN 13725 et avons utilisé le critère utilisé lors du tri rétrospectif : l'écart ΔZ . Les valeurs obtenues sont présentées en Tableau 49.

Tableau 49 : Calcul du ΔZ des différentes études tel que défini dans la norme EN13725

Etude du mélange 1		Etude du mélange 2		Etude du mélange 3		Etude du mélange 4		Etude du mélange 5	
Mélange 1	-1	Mélange 1	-1	Mélange 3	1	Mélange 3	1	Mélange 1	-1

Mélange 2	2	Mélange 2	2	Mélange 4	-11	Mélange 4	-9	Mélange 2	2
Mélange 3	-2	Mélange 5	-1	Mélange 5	2	Mélange 11	2	Mélange 3	-2
Mélange 5	1	Mélange 11	-1	Mélange 11	2	Mélange 9	3	Mélange 5	1
Mélange 10	-2			Mélange 6	2			Mélange 11	1
Mélange 11	1							Mélange 6	1
Mélange 6	1								

Ce tableau permet de quantifier l'écart de chaque concentration d'odeur à la moyenne géométrique des concentrations d'odeur d'une même étude.

La norme EN 13725 stipule qu'une valeur ayant un ΔZ en dehors de l'intervalle $[-5 ; 5]$ peut être exclue de l'ensemble des valeurs. Dans l'étude du mélange 3, il faudrait donc exclure le mélange 4 des mélanges retenus pour effectuer la régression PLS, alors qu'il s'agit du deuxième mélange le plus similaire avec le mélange 3 selon le critère de similarité défini (Tableau 50).

Tableau 50 : Critères de similarité du mélange 3 avec les autres mélanges

Echantillons	% de molécules communes (P1)	% de molécules dont AO>1% communes (P2)	Critère de similarité (P1 x P2)
Mélange 3	100	100	10000
Mélange 11	100	70	7000
Mélange 4	88	60	5280
Mélange 5	100	40	4000
Mélange 6	71	40	2840
Mélange 1	100	20	2000
Mélange 8	73	20	1460
Mélange 9	68	20	1360
Mélange 2	100	10	1000
Mélange 10	68	10	680
Mélange 7	59	10	590

Pour l'étude du mélange 4, on constate que le ΔZ du mélange 4 est en dehors de l'intervalle $[-5 ; 5]$. Cela signifie qu'il n'y aurait actuellement pas d'échantillons, dans la base de données, adaptés à l'étude statistique du mélange 4. En conséquence, les résultats obtenus doivent être considérés avec précaution.

Ces résultats nous encouragent à ne pas abaisser le critère de similarité dans l'unique but d'être en accord avec l'approche de l'expert puisque cela apporte davantage de bruit plutôt que de conduire à une représentation plus fidèle de la réalité. Le critère de similarité le plus approprié resterait donc au vu de ces résultats le critère de 2500 dans la limite de toujours conserver au moins quatre échantillons. L'utilisation de la PLS doit essentiellement s'appliquer dans une optique de complémentarité avec l'approche experte. En particulier, la PLS permettra d'observer, lors de l'examen d'un grand nombre d'échantillons, quelles molécules expliquent le plus souvent la concentration d'odeur.

4.4 Conclusion sur la démarche automatisée

En conclusion, une démarche automatisée a été mise au point qui permet de :

- Calculer automatiquement les activités d'odeur de l'échantillon ciblé et donc de retrouver rapidement les résultats de l'expert ;
- Compléter le travail de l'expert en apportant, par voie statistique, une identification complémentaire des molécules présentant un impact significatif sur la variation de la

concentration d'odeur pour un ensemble d'échantillons présentant un degré de similarité élevé. Ce dernier point nécessite un paramétrage des échantillons à croiser ensemble : il est nécessaire d'analyser des études d'au moins quatre échantillons et vérifiant un critère de similarité d'au moins 2500 tel qu'il est défini dans le chapitre.

Cette démarche automatisée a été développée à l'aide de données présentant une incertitude maîtrisée et faible en ce qui concerne les concentrations physico-chimiques et les seuils de perception.

Le chapitre suivant s'attachera à évaluer l'impact de la qualité des données d'entrée (précision des concentrations physico-chimiques et sensorielles, précision des seuils de perception) sur la qualité des résultats de la démarche.

CHAPITRE 5 : IMPACT DE LA PRECISION DES DONNEES UTILISEES SUR L'IDENTIFICATION DES MOLECULES RESPONSABLES DE L'ODEUR

5.1 Contexte

En chapitre 4, un outil automatisé permettant l'identification des molécules responsables de l'odeur d'un échantillon gazeux a été proposé. Il a été développé sur la base d'échantillons gazeux synthétiques de manière à se prémunir au mieux de l'incertitude sur les données physico-chimiques ainsi que sur les seuils de perception dont l'utilisation est essentielle pour identifier les molécules responsables de l'odeur d'un échantillon. Par ailleurs, aucune modification ou amélioration n'a été apportée au protocole de réalisation des analyses olfactométriques. Cela a permis d'évaluer la capacité d'identification intrinsèque de l'outil à identifier les molécules responsables de l'odeur.

Dans la pratique, cependant, cet outil a vocation à être utilisé sur des échantillons réels, et non synthétiques. Cela implique de gérer une incertitude souvent importante sur les données physico-chimiques (données issues de l'analyse chromatographique), les seuils de perception (données bibliographiques) et les données olfactométriques (les concentrations d'odeur). C'est pourquoi ce chapitre a pour objectif d'évaluer le niveau de précision nécessaire et suffisant sur ces différents blocs de données pour parvenir à une identification satisfaisante des molécules responsables de l'odeur.

L'incertitude sur les données physico-chimiques est principalement liée à la technique couramment utilisée pour les obtenir : la chromatographie en phase gazeuse couplée à la spectrométrie de masse (chapitre 1), en particulier à la méthode de quantification des molécules identifiées. La plus couramment utilisée est un étalonnage au toluène, mais des alternatives, associées à des niveaux de précision différents, sont envisageables.

L'incertitude sur les seuils de perception est liée, comme nous l'avons précédemment évoqué, à l'incomplétude de la base de données (le seuil de perception d'un grand nombre de molécules n'est pas renseigné dans la littérature) et à leur variabilité lorsqu'ils sont disponibles. La valeur de seuil de perception considérée comme la plus robuste est ainsi la valeur déterminée expérimentalement, dans le cadre de cette étude. Dans la pratique, cette valeur n'est que très rarement disponible, cette mesure n'étant pas systématiquement réalisée pour chaque nouvelle molécule rencontrée. Il est donc le plus souvent nécessaire de travailler avec une valeur issue de la littérature, la plus fiable possible. Parmi les différents traitements présentés précédemment, notre objectif est ici de rechercher lequel permet d'aboutir à une identification correcte des molécules responsables de l'odeur.

Enfin, l'impact de la précision de la mesure olfactométrique (mesure de la concentration d'odeur) sur l'identification des molécules responsables de l'odeur est évalué.

Ces différentes évaluations sont menées sur la base d'échantillons gazeux synthétiques décrits au chapitre 4. Chaque évaluation constitue une partie de ce chapitre.

5.2 Évaluation de l'impact de la précision de la quantification

5.2.1 Méthodologie

Dans la pratique, un mélange gazeux peut contenir plusieurs dizaines de molécules. La précision des quantifications réalisées dépend en grande partie des molécules sur lesquelles un étalonnage est réalisé.

- Lorsqu'un étalonnage est disponible pour l'une des molécules identifiées, celle-ci fait l'objet d'une quantification absolue (Ci_{abs}) exprimée en mg/m^3 .
- Lorsqu'aucun étalonnage n'est disponible pour cette molécule mais qu'elle appartient à la même famille chimique qu'une molécule pour laquelle un étalonnage a déjà été réalisé, celle-ci fait l'objet d'une semi-quantification. Sa concentration est alors exprimée en équivalent « famille chimique » ($Ci_{eq.fam}$) (« mg/m^3 équivalent famille chimique ou $mg\ eq.fam/m^3$ »). Si plusieurs étalonnages sont disponibles pour la famille chimique concernée, c'est le coefficient directeur moyen des différentes droites d'étalonnage qui est employé.
- Enfin, lorsqu'aucun étalonnage pour la famille chimique de la molécule concernée n'est disponible, celle-ci fait l'objet d'une semi-quantification en équivalent toluène ($Ci_{eq.tol}$). Sa concentration est alors exprimée en « mg/m^3 équivalent toluène ou $mg\ eq.tol/m^3$ ».

Dans la suite, ces différentes quantifications sont présentées de la moins complexe à la plus complexe à obtenir. Il s'agit ici de déterminer le niveau de précision sur la quantification nécessaire et suffisante pour obtenir une identification des molécules responsables de l'odeur satisfaisante.

Dans un premier temps, la méthode de quantification permettant de s'approcher au mieux des concentrations expérimentales (Ci_{exp}), prises comme références, est recherchée. Pour cela, les différentes quantifications sont appliquées sur deux mélanges synthétiques : les mélanges F et G. Leurs compositions respectives sont données en Annexe 26 et Annexe 27.

L'impact de la modification du niveau de précision de la quantification de ces deux mélanges sur l'identification des molécules responsables de l'odeur est ensuite évalué par voie statistique.

5.2.2 Résultats

5.2.2.1 Comparaison des méthodes de quantification

Ce paragraphe consiste à comparer différents modes de quantification avec la concentration expérimentale (C_{iexp}). Les résultats détaillés sont fournis en Annexe 26 et Annexe 27.

5.2.2.1.1 Quantification en équivalent toluène

Les résultats d'une quantification en équivalent toluène sur les deux mélanges sont comparés avec les concentrations expérimentales et illustrés par la Figure 57 et la Figure 58.

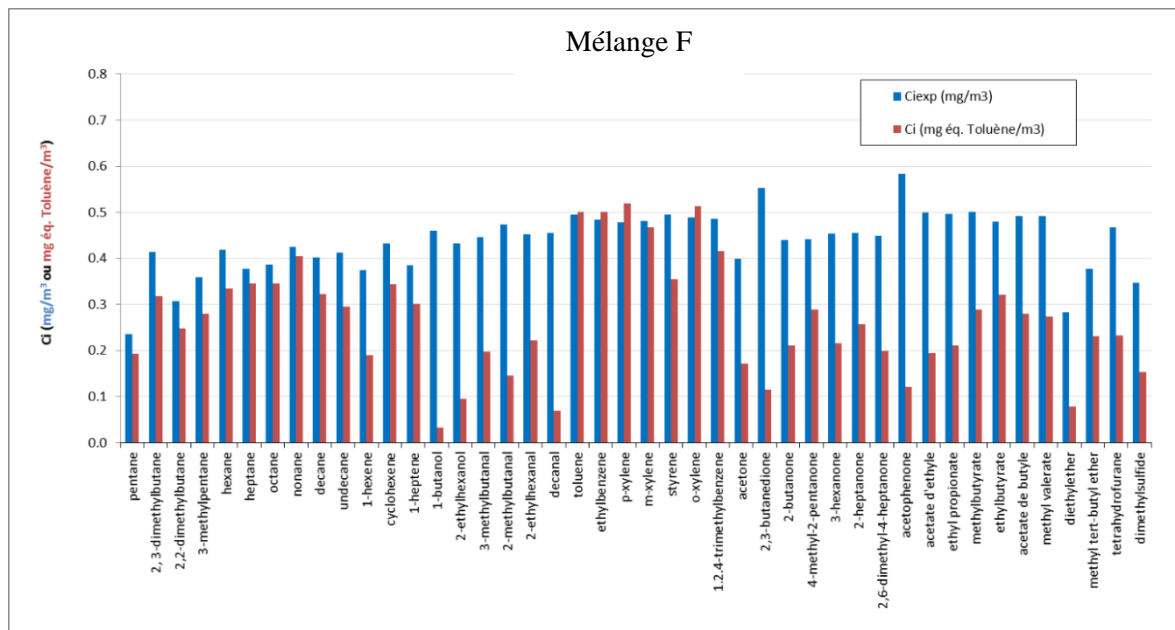


Figure 57 : Comparaison quantification en équivalent toluène / concentrations expérimentales pour le mélange F

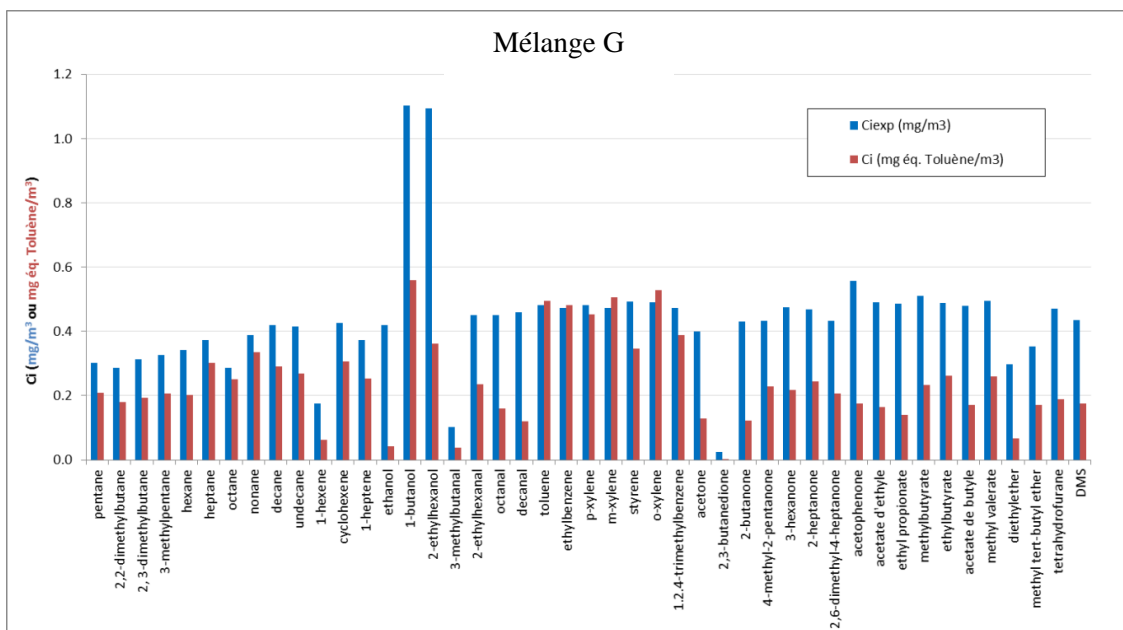


Figure 58 : Comparaison quantification en équivalent toluène / concentrations expérimentales pour le mélange G

Pour quantifier l'écart entre la valeur fournie par la quantification en équivalent toluène et la valeur expérimentale, l'erreur relative, telle qu'elle est définie dans l'Équation 9, est calculée pour chaque famille chimique (Figure 3).

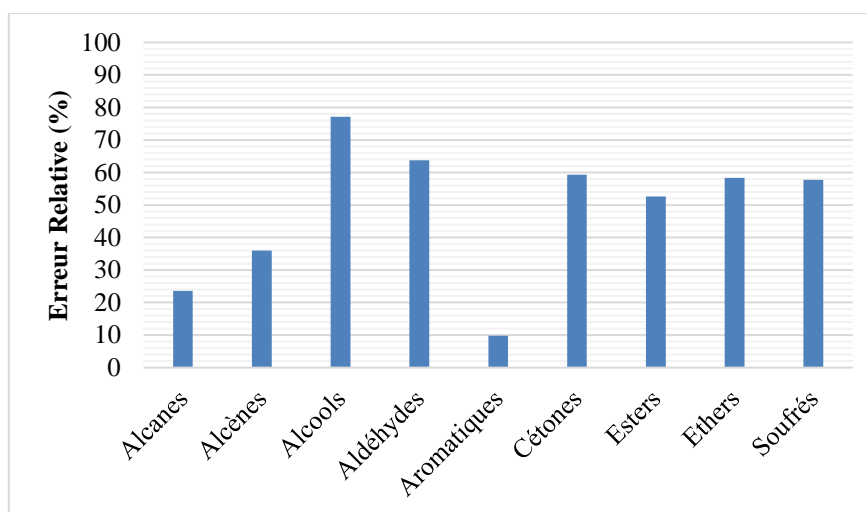


Figure 59 : Erreur relative (%) moyenne de la quantification en équivalent toluène par rapport à la concentration expérimentale sur les mélanges F et G

La première remarque qu'amène l'examen des Figure 58 et Figure 59 est que la quantification en équivalent toluène conduit à une sous-estimation systématique des concentrations des molécules présentes dans les deux mélanges, hormis pour les composés aromatiques (10%).

Pour les autres familles de molécules, la sous-estimation est soit modérée (alcane 24%, alcènes 36%) soit très importante (composés oxygénés et soufrés, environ 60%). Ces tendances sont identiques pour les deux mélanges analysés.

La quantification en équivalent toluène conduit à une sous-estimation globale des concentrations de l'ordre de 50%.

5.2.2.1.2 Quantification en équivalent famille

Les résultats d'une quantification en équivalent famille sur les deux mélanges sont illustrés par la Figure 60 et la Figure 61.

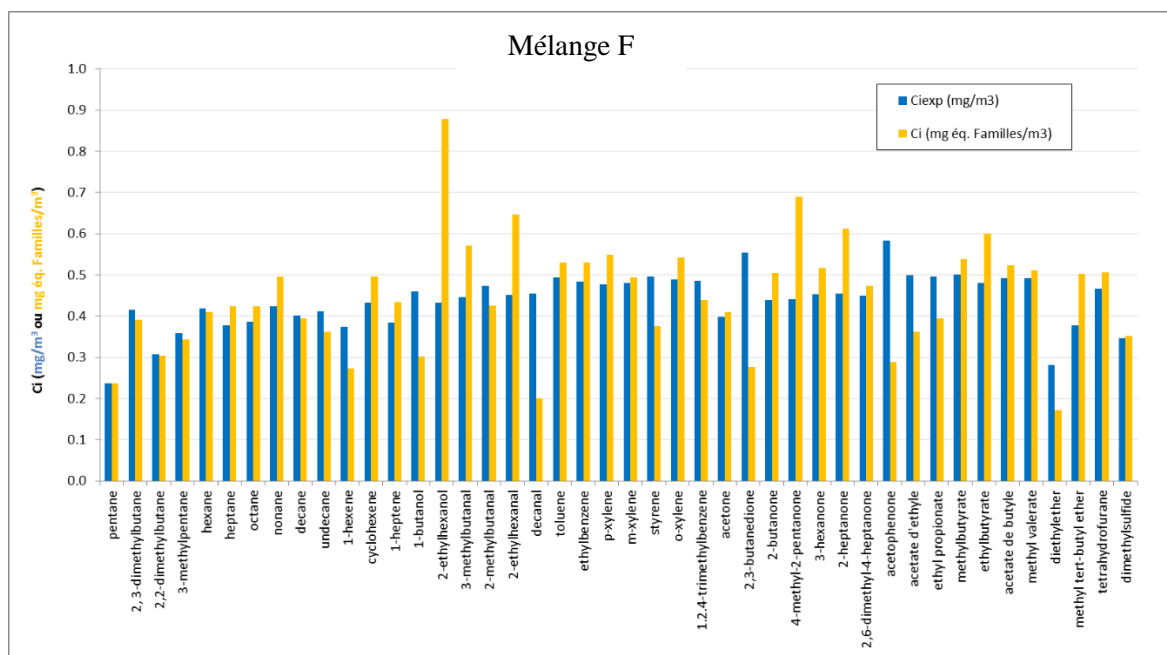


Figure 60 : Comparaison quantification en équivalent famille / concentrations expérimentales pour le mélange F

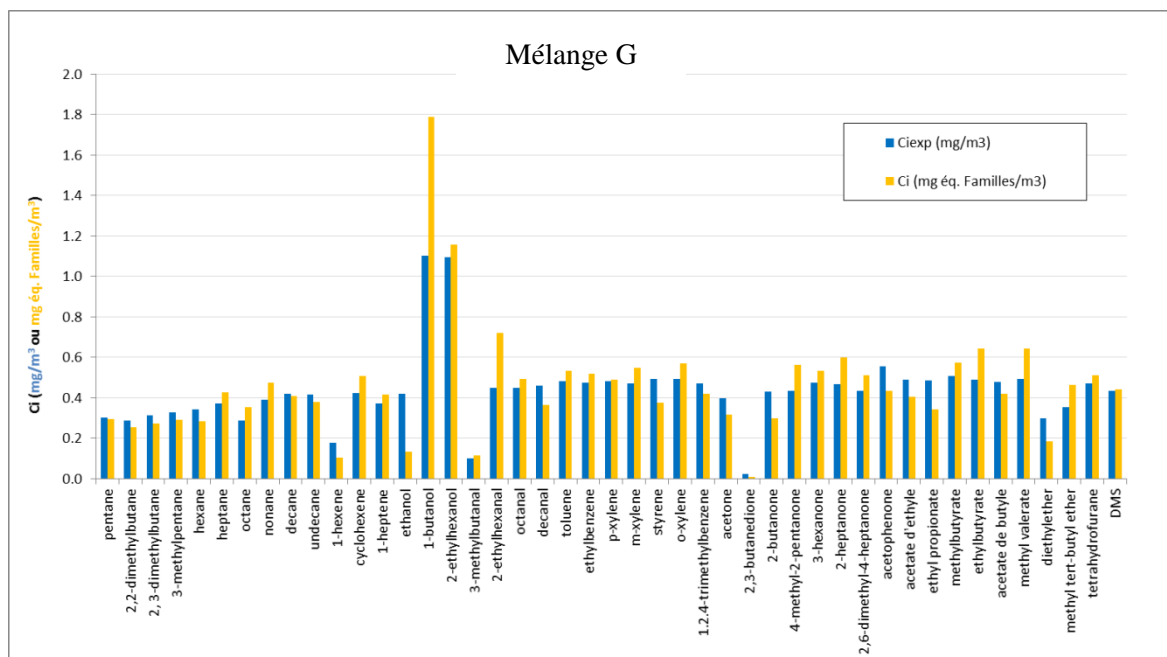


Figure 61 : Comparaison quantification en équivalent famille / concentrations expérimentales pour le mélange G

Comme précédemment, l'erreur relative entre les valeurs quantifiées et la concentration expérimentale est calculée (Figure 62).

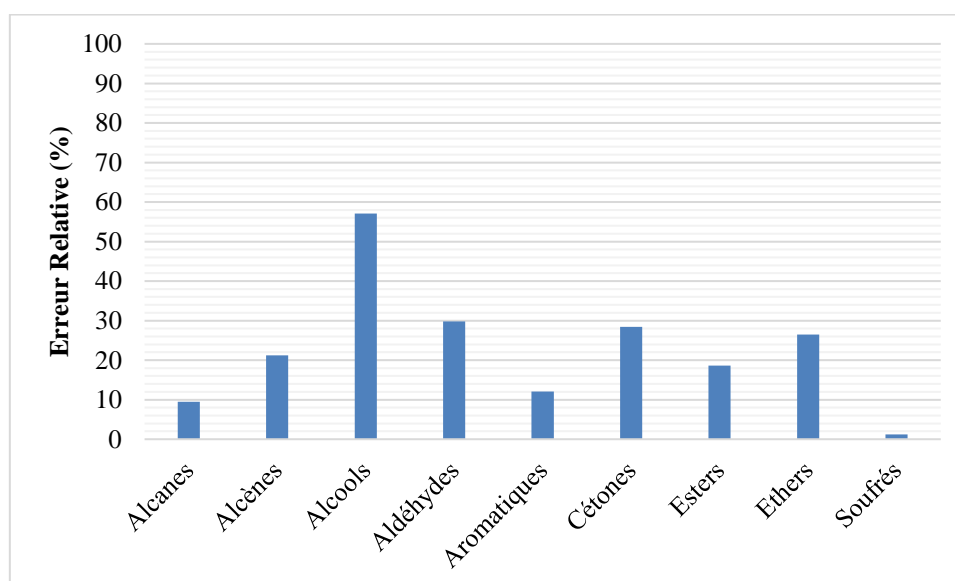


Figure 62: Erreur relative (%) moyenne de la quantification en équivalent famille par rapport à la concentration expérimentale sur les mélanges F et G

Ces deux figures permettent de constater qu'à la différence de la quantification en équivalent toluène, la quantification en équivalent famille permet de cerner avec beaucoup plus de précision les concentrations expérimentales (Erreur relative moyenne de 23%). Par ailleurs, cette quantification ne conduit pas systématiquement à une sous-estimation des concentrations.

5.2.2.1.3 Quantification absolue

Les résultats d'une quantification absolue sur les deux mélanges sont illustrés par la Figure 63 et la Figure 64.

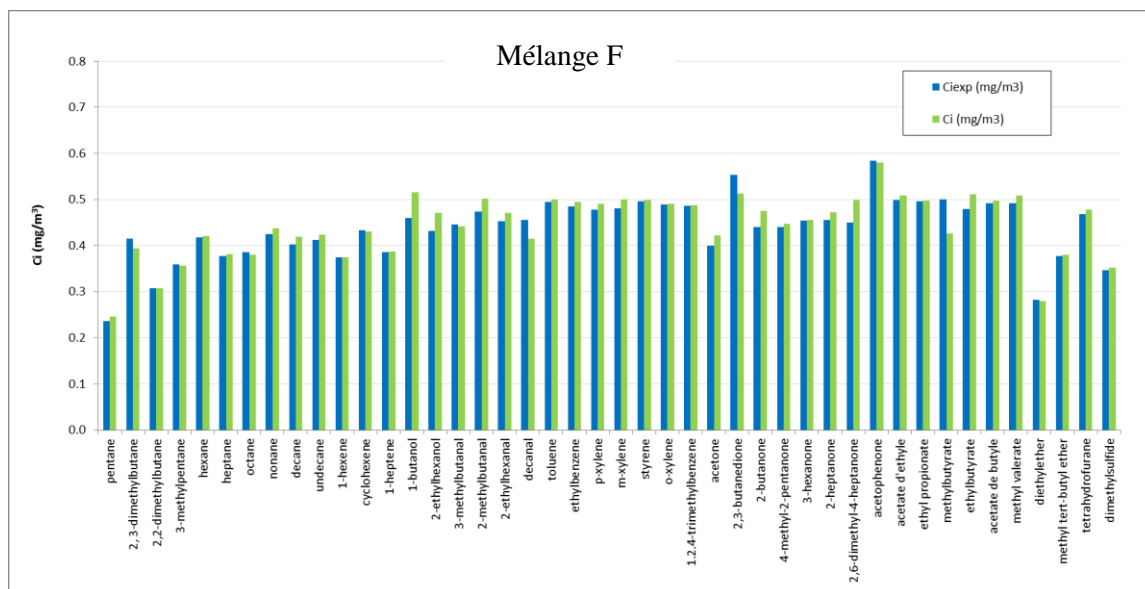


Figure 63 : Comparaison quantification absolue / concentrations expérimentales pour le mélange F

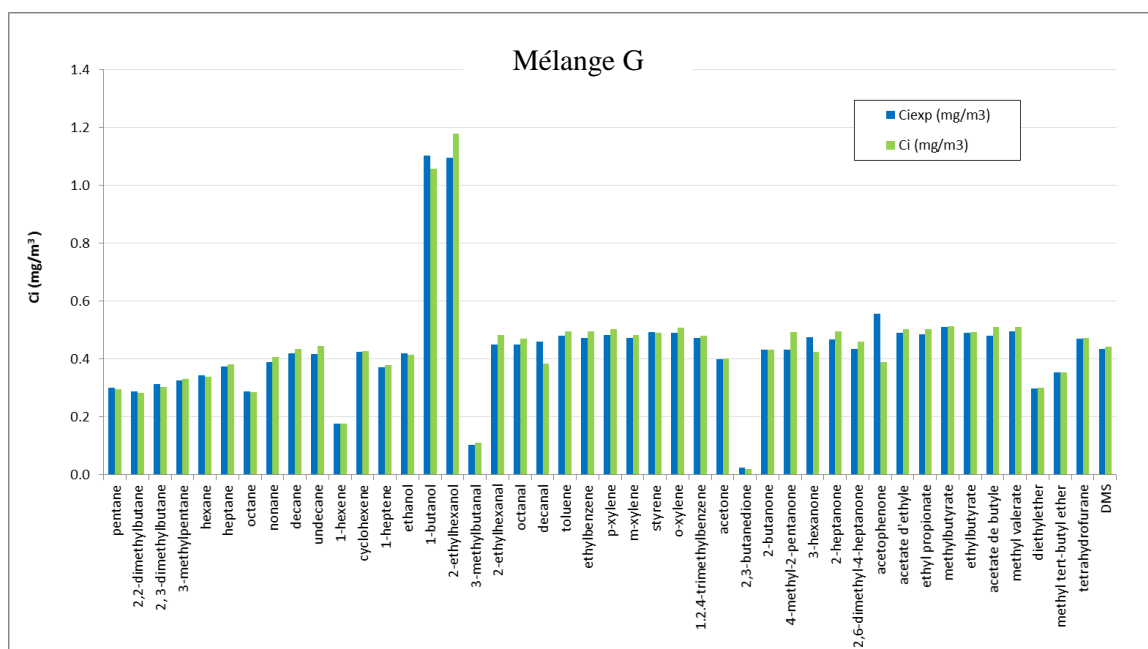


Figure 64 : Comparaison quantification absolue / concentrations expérimentales pour le mélange G

Ces figures illustrent l'excellente concordance entre les valeurs de concentrations obtenues par quantification absolue et les concentrations expérimentales. De nouveau, l'erreur relative moyenne entre les valeurs quantifiées et la concentration expérimentale est présentée en Figure 65.

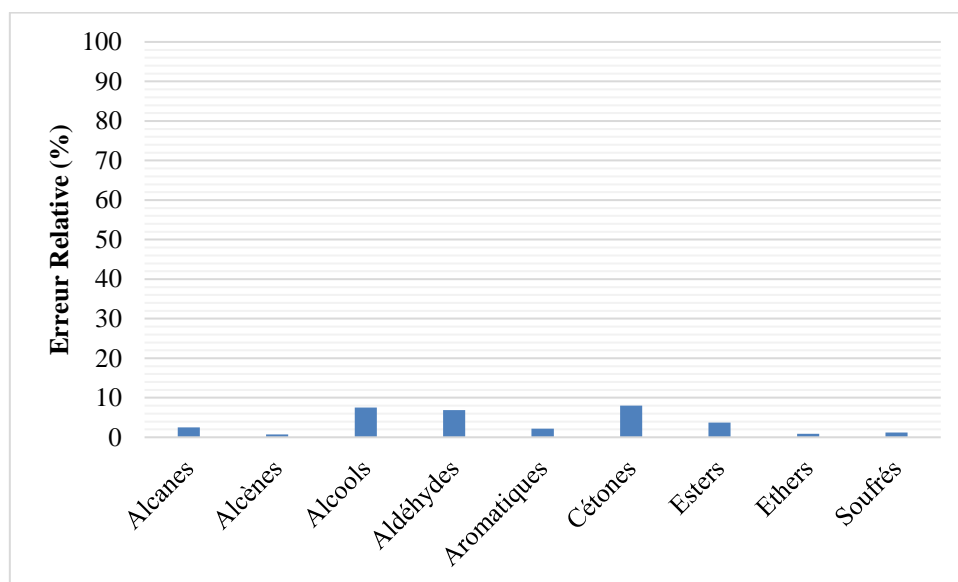


Figure 65 : Erreur relative (%) de la quantification absolue par rapport à la concentration expérimentale sur les mélanges F et G

L'erreur relative moyenne est dans ce cas inférieure à 4%.

5.2.2.1.4 Bilan sur les méthodes de quantification

À l'issue de cette partie de l'étude, il apparaît clairement que les trois modes de quantification utilisés se soldent par des niveaux de précision très différents. Le Tableau 51 ci-dessous les synthétise, par famille et au global.

Tableau 51 : Pourcentages d'erreur relative par rapport à la concentration expérimentale

	Équivalent-Toluène	Équivalent-Famille	Quantification absolue
Alcanes	24	9	3
Alcènes	36	21	1
Alcools	77	57	7
Aldéhydes	64	30	7
Aromatiques	10	12	2
Cétones	59	28	8
Esters	53	19	4
Éthers	58	26	1
Soufrés	58	1	1
Moyenne	49	23	4

Ces résultats montrent que les pourcentages d'erreur les plus importants sont obtenus pour les composés oxygénés (alcools, aldéhydes, cétones). Ce constat peut s'expliquer par leur polarité et par les conditions

analytiques utilisées (colonne apolaire) qui peuvent altérer la forme des pics chromatographiques, et par voie de conséquence, les conditions d'intégration.

5.2.2.2 Impact du mode de quantification sur l'identification des molécules responsables de l'odeur

Les molécules identifiées comme responsables de l'odeur et les activités d'odeur associées sont présentées en Tableau 52 pour les mélanges F et G à l'aide des graphes de l'Annexe 28 et de l'Annexe 29. L'approche statistique y est comparée à l'approche experte, prise comme référence. Pour réaliser cette identification, les valeurs des seuils de perception utilisées pour le calcul des activités d'odeur des molécules sont les mêmes, quel que soit le mode de quantification ou l'approche utilisés.

Tableau 52 : Molécules identifiées comme responsables de la concentration d'odeur suivant le mode de quantification et l'approche (mélanges F et G)

		Mélange F (CO = 7050 UO _E /m ³)							
		Concentration théorique		Quantification absolue		Quantification équivalent famille		Quantification équivalent toluène	
		Molécule	OAV _{théo}	Molécule	OAV _{Abs}	Molécule	OAV _{Eq,Fam}	Molécule	OAV _{Eq,Tol}
Approche expert	butane-2,3-dione	4683	butane-2,3-dione	4360	3-methylbutanal	2601	butane-2,3-dione	986	
	3-methylbutanal	2046	3-methylbutanal	2008	butane-2,3-dione	2356	3-methylbutanal	895	
	ethyl butanoate	830	ethyl butanoate	883	ethyl butanoate	1037	ethyl butanoate	556	
	hex-1-ene	708	hex-1-ene	716	hex-1-ene	523	hex-1-ene	363	
Approche statistique	2-methylbutanal	280	2-methylbutanal	299	2-methylbutanal	253	2-methylbutanal	87	
	ethyl butanoate	830	dimethylsulfide	88	dimethylsulfide	88	ethyl butanoate	555	
	butane-2,3-dione	4683	butane-2,3-dione	4360	3-methylbutanal	2601	3-methylbutanal	895	
	3-methylbutanal	2046	3-methylbutanal	2008	hex-1-ene	523	hex-1-ene	363	
	hex-1-ene	708	ethyl butanoate	883	ethyl butanoate	1037			
			hex-1-ene	716					

		Mélange G (CO = 5493 UO _E /m ³)							
		Concentration théorique		Quantification absolue		Quantification équivalent famille		Quantification équivalent toluène	
		Molécule	OAV _{théo}	Molécule	OAV _{Abs}	Molécule	OAV _{Eq,Fam}	Molécule	OAV _{Eq,Tol}
Approche expert	octanal	900	octanal	943	ethyl butanoate	1116	ethyl butanoate	452	
	ethyl butanoate	847	ethyl butanoate	851	octanal	986	octanal	322	
	3-methylbutanal	455	3-methylbutanal	499	3-methylbutanal	518	3-methylbutanal	169	
	hex-1-ene	345	hex-1-ene	340	hex-1-ene	198	hex-1-ene	120	
	butane-2,3-dione	170	butane-2,3-dione	163					
Approche statistique	2-methylbutanal	72	ethyl butanoate	851	ethyl butanoate	1116	ethyl butanoate	452	
	butane-2,3-dione	170	butane-2,3-dione	163	methyl valerate	64	3-methylbutanal	169	
	dimethylsulfide	110	3-methylbutanal	499	hex-1-ene	198	methyl valerate	26	
	methyl valerate	49	hex-1-ene	340	3-methylbutanal	518	acetophenone	14	
	acetophenone	43	methyl valerate	51	butane-2,3-dione	75	hex-1-ene	120	
	3-methylbutanal	455							
	ethyl butanoate	847							
hex-1-ene	345								

Concernant l'approche expert, l'examen de ces résultats permet de faire ressortir plusieurs constats :

- Quel que soit le mélange considéré, les molécules identifiées sont quasiment les mêmes. Seule la quantification absolue dans le cas du mélange G permet de faire ressortir la butane 2-3 dione.
- L'approche de l'expert classe les molécules par ordre d'importance dans l'explication de la concentration d'odeur du mélange étudié.
- L'amélioration de la quantification fait tendre la somme des activités d'odeur des molécules identifiées vers la concentration d'odeur mesurée. On peut considérer, en première analyse, que cela traduit le fait que les molécules qui ressortent de l'analyse de l'expert sont bien à même d'expliquer l'odeur du mélange considéré.
- Si l'on considère que les activités d'odeur théoriques sont représentatives des poids réels des molécules dans la concentration d'odeur du mélange, on constate que l'amélioration de la quantification permet de se rapprocher de cette référence.

Concernant l'approche statistique, plusieurs points peuvent être soulignés :

- Cette approche classe les molécules par ordre de VIP, ce qui ne correspond pas à l'ordre des activités d'odeur des molécules du mélange étudié.
- Des molécules de moindre poids dans l'odeur apparaissent dans l'identification : le 2-methylbutanal et le dimethylsulfide pour les deux mélanges, le méthyl-valérate et l'acétophénone pour le mélange G.
- Des molécules identifiées par l'expert comme ayant un poids élevé dans l'explication de la concentration d'odeur n'apparaissent pas dans les résultats de l'approche statistique : l'octanal pour le mélange G, quel que soit le mode de quantification et la butane-2,3-dione pour le mélange F, lors de l'application de la quantification en équivalent famille et en équivalent toluène.

5.2.3 Bilan sur l'impact du mode de quantification

La différence entre les résultats des deux approches peut s'expliquer par le fait que la PLS est bâtie à partir d'un ensemble de données constitué par les résultats d'analyse de mélanges présentant des taux de similarité (en termes de molécules présentes et de concentrations – chapitre 4) certes élevés, mais au sein desquels des différences existent quant aux molécules responsables de l'odeur. Cette technique fait donc ressortir des molécules expliquant au mieux la variation de la concentration d'odeur de l'ensemble des données utilisées, et non spécifiquement les molécules responsables de l'odeur du mélange étudié.

Il résulte logiquement des constatations précédentes que l'impact de l'amélioration de la quantification sur l'identification des molécules responsables de l'odeur est moindre dans le cas de l'approche statistique, par le fait même qu'elle peut conduire à ne pas considérer une ou plusieurs molécules importantes (Octanal dans le cas du mélange 5).

5.3 Évaluation de l'impact de la précision des valeurs de seuils de perception

5.3.1 Méthodologie

L'évaluation a été réalisée sur la base de 5 mélanges gazeux synthétiques. Cela permet d'utiliser les concentrations expérimentales, et de n'observer ainsi qu'uniquement l'impact du niveau de précision de la valeur du seuil de perception utilisé. Ces valeurs de seuils de perception sont présentées suivant leur degré de complexité d'obtention croissant.

La solution la plus simple pour attribuer un seuil de perception à une molécule donnée est, comme présenté au chapitre 1, d'utiliser la moyenne géométrique des seuils de perception des molécules de la famille chimique à laquelle appartient cette molécule. Cela nécessite d'inventorier des valeurs de seuils de perception disponibles dans la littérature pour les différentes familles chimiques. Il s'agit du « scénario 1 ».

Dans le cas où différentes valeurs de seuil de perception sont disponibles dans la littérature pour la molécule considérée, on peut lui attribuer le mode³ de ces valeurs de seuils de perception. Si une seule valeur de seuil est disponible, elle est utilisée telle quelle. Enfin, si aucune valeur de seuil n'est disponible pour la molécule considérée, on lui attribue la moyenne géométrique des seuils de la famille chimique à laquelle elle appartient. C'est le « scénario 2 ».

Pour répondre à la problématique de la variabilité des valeurs de seuils de perception disponibles, une méthode de fiabilisation est proposée dans le « scénario 3 ». La démarche utilisée dans ce scénario est identique à la précédente, si ce n'est que la base de données utilisée est la base de données fiabilisée.

Une modélisation prédictive globale à partir des seuils de perception disponibles permettant de pallier l'incomplétude des données fait l'objet du « scénario 4 » et une modélisation prédictive locale permettant d'affiner le modèle global par une modélisation locale de l'erreur, celui du « scénario 5 ». Dans ces deux scénarios, ce sont exclusivement les valeurs modélisées qui sont utilisées.

³ Valeur d'un caractère quantitatif discret ayant la fréquence la plus élevée (LAROUSSE 2018)

Enfin, l'approche de l'expert décrite au chapitre 4, qui consiste à utiliser en priorité le seuil de perception mesuré par ses soins, puis, par défaut le seuil disponible (ou le mode des seuils disponibles) pour la molécule, et enfin, si aucune valeur n'est disponible, la moyenne géométrique des seuils de la famille chimique à laquelle appartient la molécule constitue le « scénario 6 ».

Afin de pouvoir évaluer l'apport de la démarche de fiabilisation des seuils de perception, celle-ci n'a jamais été utilisée par l'expert.

Un logigramme d'utilisation des seuils de perception en fonction des disponibilités est présenté en Tableau 53.

Tableau 53 : Logigramme d'utilisation des seuils de perception

	SP expérimental	modélisation		SP littérature			
		globale	affinée	BDD non fiabilisée		BDD fiabilisée	
				Moyenne géométrique des valeurs disponibles pour la molécule	Moyenne géométrique des valeurs disponibles pour la famille de la molécule	Moyenne géométrique des valeurs disponibles pour la molécule	Moyenne géométrique des valeurs disponibles pour la famille de la molécule
Scénario 1					1		
Scénario 2				1	2		
Scénario 3						1	2
Scénario 4		1					
Scénario 5			1				
Scénario 6	1			2	3		

Cette partie consiste à trouver parmi les 5 scénarios, celui permettant de s'approcher au mieux des résultats de l'expert et répondant le plus au principe de parcimonie. Autrement dit, il s'agit d'identifier le scénario le moins complexe à mettre en œuvre et permettant néanmoins d'obtenir le résultat le plus satisfaisant.

Le choix du scénario le plus adapté est basé sur deux principes. Le premier consiste à calculer la somme des activités d'odeur individuelles pour chaque mélange et chaque scénario de manière à obtenir la « concentration d'odeur théorique » dans chacun des cas correspondants. Cela permet d'identifier le scénario permettant le plus souvent d'approcher la concentration d'odeur mesurée par olfactométrie. Le second consiste à appliquer l'outil statistique présenté au chapitre 4 et à comparer le résultat de l'identification des molécules responsables de l'odeur aux résultats de l'expert. Cela permet d'identifier le scénario le plus adapté.

5.3.2 Résultats

5.3.2.1 Identification du scénario permettant la meilleure estimation de la concentration d'odeur des mélanges

Les valeurs de seuils de perception obtenues grâce aux six scénarios sont présentées en Tableau 54 en fonction du degré de complexité de leur obtention. Un code couleur est attribué par ordre de grandeur.

Tableau 54 : Valeurs des seuils de perception obtenues grâce aux différents scénarios

		Ordre de grandeur							
		-5	-4	-3	-2	-1	0	1	2
No_CAS	Molécules	Scénario 1	Scénario 2	Scénario 3	Scénario 4	Scénario 5	Scénario 6		
109-66-0	pentane	1.36E+01	4.10E+00	4.10E+00	2.27E+00	2.62E+00	4.10E+00		
110-54-3	hexane	1.36E+01	5.30E+00	5.31E+00	2.15E+00	3.11E+00	1.24E+01		
111-65-9	octane	1.36E+01	3.55E+01	9.89E+00	2.29E+00	3.19E+00	1.22E+01		
111-84-2	nonane	1.36E+01	1.20E+01	1.20E+01	2.14E+00	3.01E+00	1.20E+01		
1120-21-4	undecane	1.36E+01	9.59E+00	9.59E+00	9.38E-01	9.27E-01	9.59E+00		
124-18-5	decane	1.36E+01	1.60E+02	3.60E+00	1.26E+00	5.28E+00	1.60E+02		
142-82-5	heptane	1.36E+01	5.71E+01	2.70E+00	2.40E+00	3.28E+00	5.71E+01		
75-83-2	2,2-dimethylbutane	1.36E+01	7.00E+01	7.00E+01	1.53E+00	2.21E+01	7.00E+01		
79-29-8	2, 3-dimethylbutane	1.36E+01	1.50E+00	1.50E+00	1.58E+00	2.28E+00	1.50E+00		
96-14-0	3-methylpentane	1.36E+01	3.10E+01	3.10E+01	1.53E+00	3.06E+00	3.10E+01		
110-83-8	cyclohexene	8.31E-01	6.00E-01	5.99E-01	4.56E-01	3.73E-01	2.35E-01		
111-66-0	1-octene	8.31E-01	1.85E+00	5.75E-03	1.75E-01	9.04E-03	7.18E-03		
592-41-6	1-hexene	8.31E-01	4.80E-01	4.79E-01	1.81E-01	7.00E-04	5.22E-04		
592-76-7	1-heptene	8.31E-01	1.50E+00	4.66E-01	1.46E-01	2.26E-01	1.43E-01		
104-76-7	2-ethylhexanol	1.22E+00	7.99E-01	4.00E-01	2.02E-01	2.40E-01	3.81E-01		
64-17-5	ethanol	1.22E+00	1.87E+01	1.07E+00	9.49E-01	1.18E-01	9.36E-02		
71-36-3	1-butanol	1.22E+00	4.21E-01	1.35E-01	1.76E-01	2.07E-01	5.21E-02		
112-31-2	decanal	2.10E-02	5.67E-02	2.03E-02	1.62E-02	1.00E-02	2.81E-03		
123-05-7	2-ethylhexanal	2.10E-02	2.10E-02	7.73E-02	2.84E-02	4.88E-02	7.73E-02		
124-13-0	octanal	2.10E-02	7.04E-03	2.63E-03	2.38E-02	1.05E-02	5.00E-04		
590-86-3	3-methylbutanal	2.10E-02	7.50E-04	3.50E-04	5.46E-03	2.77E-04	2.20E-04		
96-17-3	2-methylbutanal	2.10E-02	2.10E-02	1.68E-03	5.16E-03	2.66E-03	1.68E-03		
100-41-4	ethylbenzene	1.63E-01	3.16E-01	1.41E-01	1.57E-01	1.66E-01	3.16E-01		
100-42-5	styrene	1.63E-01	1.20E-01	1.37E-01	7.45E-02	7.57E-02	5.92E-02		
106-42-3	p-xylene	1.63E-01	5.79E-01	3.61E-01	1.97E-01	2.28E-01	5.79E-01		
108-38-3	m-xylene	1.63E-01	5.08E-01	4.86E-01	1.98E-01	2.26E-01	5.08E-01		
108-88-3	toluene	1.63E-01	2.43E+00	1.17E+00	4.22E-01	4.08E-01	1.22E+00		
95-47-6	o-xylene	1.63E-01	1.44E+00	1.48E+00	1.89E-01	2.38E-01	1.49E+00		
95-63-6	1.2.4-trimethylbenzene	1.63E-01	6.43E-01	6.43E-01	1.15E-01	1.62E-01	6.43E-01		
108-10-1	4-methyl-2-pentanone	5.68E-01	7.80E-01	7.17E-01	1.01E-01	1.89E-01	7.21E-01		
108-83-8	2,6-dimethyl-4-heptanone	5.68E-01	9.28E+00	6.00E-01	1.91E-01	2.53E-01	4.01E-01		
110-43-0	2-heptanone	5.68E-01	2.60E-01	4.80E-02	1.09E-01	2.23E-01	7.81E-02		
431-03-8	2,3-butanedione (diacetyl)	5.68E-01	5.74E-04	6.53E-04	1.07E-01	3.71E-04	1.17E-04		
589-38-8	3-hexanone	5.68E-01	5.68E-01	1.09E-01	9.79E-02	1.48E-01	1.09E-01		
591-78-6	2-hexanone	5.68E-01	7.08E-01	3.59E-01	1.10E-01	2.72E-01	6.03E-01		
67-64-1	acetone	5.68E-01	5.56E+01	4.91E+00	1.28E+00	1.95E+00	7.75E+00		
78-93-3	2-butanone	5.68E-01	5.59E+00	2.35E+00	5.15E-01	1.05E+00	2.54E+00		
98-86-2	acetophenone	5.68E-01	1.39E-01	1.30E-02	1.76E-02	1.80E-02	1.30E-02		
105-37-3	ethyl propionate	2.30E-01	1.20E+00	2.52E-02	6.97E-02	8.73E-02	2.19E-02		
105-54-4	ethylbutyrate	2.30E-01	1.70E-05	8.47E-05	3.46E-02	1.40E-03	5.78E-04		
123-86-4	acetate de butyle	2.30E-01	1.00E-02	5.36E-02	4.34E-02	3.64E-02	5.13E-02		
141-78-6	acetate d'ethyl	2.30E-01	3.31E+00	3.28E+00	3.31E-01	7.31E-01	2.91E+00		
623-42-7	methylbutyrate	2.30E-01	7.69E-02	3.16E-02	7.89E-02	5.26E-02	3.32E-02		
624-24-8	methyl valerate	2.30E-01	1.10E-02	1.05E-02	3.71E-02	4.01E-02	1.01E-02		
109-99-9	tetrahydrofurane	1.64E+00	8.99E+01	2.26E+00	5.60E-01	6.73E-01	2.26E+00		
1634-04-4	MTBE	1.64E+00	6.29E-01	2.50E-02	1.18E-01	9.95E-02	2.50E-02		
60-29-7	diethylether	1.64E+00	7.70E-01	2.24E-01	1.88E-01	1.59E-01	1.00E-01		
75-18-3	dimethylsulfide(DMS)	1.84E-03	5.61E-03	4.68E-03	5.04E-03	4.02E-03	4.00E-03		

Ce tableau permet d'observer que les seuils de perceptions obtenus pour certaines molécules sont du même ordre de grandeur quel que soit le scénario utilisé (4-méthyl-2-pentanone ou diméthylsulfure), tandis qu'ils varient beaucoup pour d'autres (2,3-butanedione).

De plus, les données issues du scénario 5 sont fréquemment les plus proches de celles issues du scénario 6, qui correspond à la démarche de l'expert. La limite de ce résultat réside dans le fait que les scénarios 4 et 5 utilisent les valeurs de seuils de perception mesurées au laboratoire comme données de construction. Or ces valeurs sont elles-mêmes majoritairement utilisées dans le scénario 6. Cela peut expliquer la proximité entre les résultats des scénarios 5 et 6. Cependant, comme cela a pu être montré dans le chapitre 3, il n'a pas été noté de sur-apprentissage du modèle global sinon un sur-apprentissage de l'affinage qui ne dégrade pas pour autant la valeur de validation. Cela encourage donc à utiliser le scénario 5.

À partir des concentrations expérimentales présentées en Tableau 38 du chapitre 4, la « concentration d'odeur théorique » (somme des activités d'odeur) obtenue pour chaque mélange et pour chaque scénario est présentée en Tableau 55 et illustrée en Figure 66.

Tableau 55 : Concentrations d'odeur expérimentales (CO_{exp}) et théoriques (CO_{th})

	CO_{exp} (UO/ m^3)	CO_{th} Scénario 1 (UO/ m^3)	CO_{th} Scénario 2 (UO/ m^3)	CO_{th} Scénario 3 (UO/ m^3)	CO_{th} Scénario 4 (UO/ m^3)	CO_{th} Scénario 5 (UO/ m^3)	CO_{th} Scénario 6 (UO/ m^3)
Mélange 1	13798	1014	88917	22019	1235	8225	12113
Mélange 2	31175	1603	150104	42135	2208	22304	45389
Mélange 3	9950	2196	8140	14935	8385	8006	15738
Mélange 4	788	325	561	596	738	820	1017
Mélange 5	17653	1785	146355	33667	1734	7306	16918

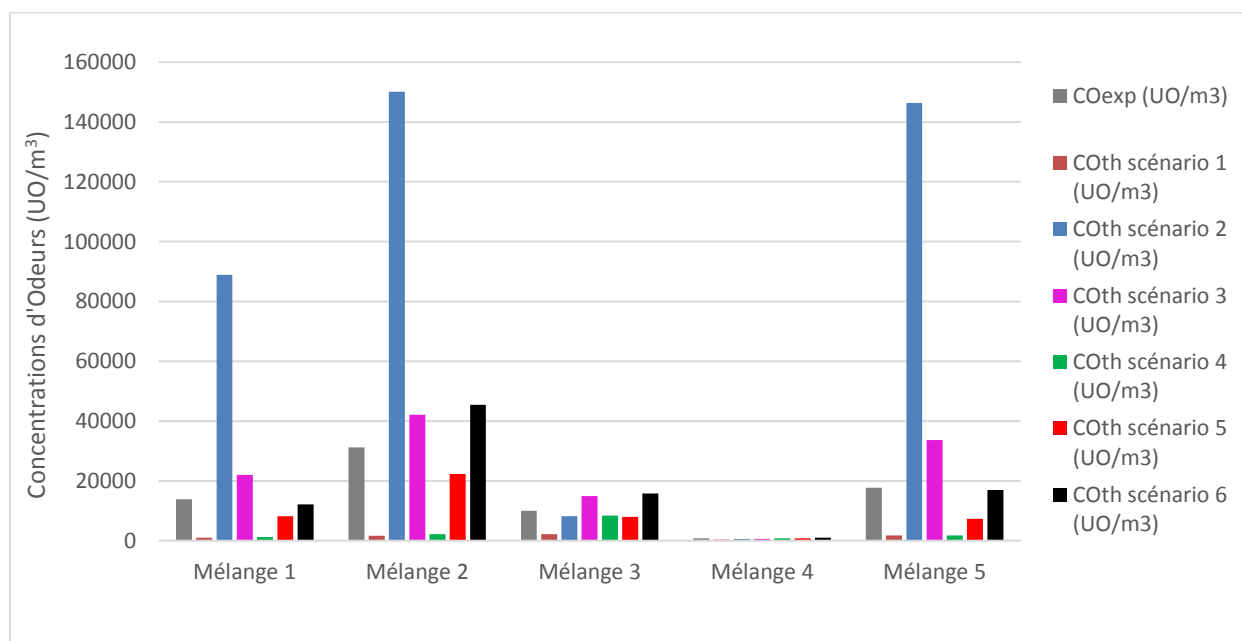


Figure 66 : Concentrations d'odeur expérimentales (CO_{exp}) et théoriques (CO_{th})

Sur la base des valeurs des concentrations d'odeur théoriques et de la concentration d'odeur expérimentale (Tableau 55) il est possible de calculer la valeur de l'erreur relative $\left(\left| \frac{CO_{th} - CO_{exp}}{CO_{exp}} \right| \right)$ et sa valeur moyenne par mélange ou par scénario (Tableau 56).

Tableau 56 : Erreurs relatives pour les différents mélanges et scénarios

Mélanges	Erreurs Relatives (%)						Erreur Relative Moyenne (%)
	Scénario 1	Scénario 2	Scénario 3	Scénario 4	Scénario 5	Scénario 6	
1	93	544	60	91	40	12	140
2	95	381	35	93	28	46	113
3	78	18	50	16	20	58	40
4	59	29	24	6	4	29	25
5	90	729	91	90	59	4	177
Erreur Relative Moyenne (%)	83	340	52	59	30	30	

L'examen des pourcentages d'erreur relative moyens par scénario confirme que les différents traitements de fiabilisation des seuils de perception permettent d'améliorer l'estimation de la concentration d'odeur expérimentale. Il confirme aussi que le scénario 5 est celui qui permet d'obtenir les résultats les plus proches du scénario 6 (approche expert).

Des niveaux d'erreur relative sensiblement différents sont observés pour les mélanges 3 et 4. Pour l'expliquer, un parallèle peut être fait entre la concentration des molécules dans les mélanges et la valeur du seuil de perception selon le scénario considéré (Tableau 57).

Tableau 57 : Parallèle entre la concentration des molécules dans les mélanges et valeur du seuil de perception selon le scénario

No_CAS	Molécules	Mélange	Mélange	Mélange	Mélange	Mélange	Scénario	Scénario	Scénario	Scénario	Scénario	Scénario
		1 (mg/m ³)	2 (mg/m ³)	3 (mg/m ³)	4 (mg/m ³)	5 (mg/m ³)	1 (mg/m ³)	2 (mg/m ³)	3 (mg/m ³)	4 (mg/m ³)	5 (mg/m ³)	6 (mg/m ³)
109-66-0	pentane	1.62	1.18	5.28	1.03	1.51	1.36E+01	4.10E+00	4.10E+00	2.27E+00	2.62E+00	4.10E+00
110-54-3	hexane	1.42	2.09	7.68	1.55	1.71	1.36E+01	5.30E+00	5.31E+00	2.15E+00	3.11E+00	1.24E+01
111-65-9	octane	6.47	1.93	8.91	1.9	1.44	1.36E+01	3.55E+01	9.89E+00	2.29E+00	3.19E+00	1.22E+01
111-84-2	nonane	4.85	2.12	9.26	1.99	1.95	1.36E+01	1.20E+01	1.20E+01	2.14E+00	3.01E+00	1.20E+01
1120-21-4	undecane	1.73	2.06	9.36	2.07	2.08	1.36E+01	9.59E+00	9.59E+00	9.38E-01	9.27E-01	9.59E+00
124-18-5	decane	1.84	2.01	10.66	2.14	2.1	1.36E+01	1.60E+02	3.60E+00	1.26E+00	5.28E+00	1.60E+02
142-82-5	heptane	4.47	1.89	9	2.01	1.87	1.36E+01	5.71E+01	2.70E+00	2.40E+00	3.28E+00	5.71E+01
75-83-2	2,2-dimethylbutane	5.02	1.54	7.6	1.65	1.43	1.36E+01	7.00E+01	7.00E+01	1.53E+00	2.21E+01	7.00E+01
79-29-8	2,3-dimethylbutane	1.8	2.07	7.39	1.32	1.56	1.36E+01	1.50E+00	1.50E+00	1.58E+00	2.28E+00	1.50E+00
96-14-0	3-methylpentane	5.39	1.8	6.33	1.78	1.63	1.36E+01	3.10E+01	3.10E+01	1.53E+00	3.06E+00	3.10E+01
110-83-8	cyclohexene	1.6	2.16	10.26	1.03	2.13	8.31E-01	6.00E-01	5.99E-01	4.56E-01	3.73E-01	2.35E-01
111-66-0	1-octene	1.77	1.99	0.97	0	1.95	8.31E-01	1.85E+00	5.75E-03	1.75E-01	9.04E-03	7.18E-03
592-41-6	1-hexene	1.72	1.87	0	0	0.88	8.31E-01	4.80E-01	4.79E-01	1.81E-01	7.00E-04	5.22E-04
592-76-7	1-heptene	1.81	1.93	8.58	0.93	1.86	8.31E-01	1.50E+00	4.66E-01	1.46E-01	2.26E-01	1.43E-01
104-76-7	2-ethylhexanol	1.55	2.16	10.39	1.61	5.47	1.22E+00	7.99E-01	4.00E-01	2.02E-01	2.40E-01	3.81E-01
64-17-5	ethanol	0	0	0	19.94	2.1	1.22E+00	1.87E+01	1.07E+00	9.49E-01	1.18E-01	9.36E-02
71-36-3	1-butanol	1.15	2.3	10.42	1.13	5.52	1.22E+00	4.21E-01	1.35E-01	1.76E-01	2.07E-01	5.21E-02
112-31-2	decanal	0.56	2.27	1	0	2.29	2.10E-02	5.67E-02	2.03E-02	1.62E-02	1.00E-02	2.81E-03

123-05-7	2-ethylhexanal	1.1	2.26	10.64	1.12	2.26	2.10E-02	2.10E-02	7.73E-02	2.84E-02	4.88E-02	7.73E-02
124-13-0	octanal	0	0	0	0	2.25	2.10E-02	7.04E-03	2.63E-03	2.38E-02	1.05E-02	5.00E-04
590-86-3	3-methylbutanal	0.9	2.23	0	0	0.51	2.10E-02	7.50E-04	3.50E-04	5.46E-03	2.77E-04	2.20E-04
96-17-3	2-methylbutanal	1.4	2.37	0	0	0.62	2.10E-02	2.10E-02	1.68E-03	5.16E-03	2.66E-03	1.68E-03
100-41-4	ethylbenzene	1.67	2.42	11.2	1.2	2.37	1.63E-01	3.16E-01	1.41E-01	1.57E-01	1.66E-01	3.16E-01
100-42-5	styrene	1.73	2.48	11.56	23.95	2.46	1.63E-01	1.20E-01	1.37E-01	7.45E-02	7.57E-02	5.92E-02
106-42-3	p-xylene	1.38	2.39	11.02	1.14	2.41	1.63E-01	5.79E-01	3.61E-01	1.97E-01	2.28E-01	5.79E-01
108-38-3	m-xylene	1.74	2.4	11.24	1.24	2.36	1.63E-01	5.08E-01	4.86E-01	1.98E-01	2.26E-01	5.08E-01
108-88-3	toluene	1.86	2.47	10.91	1.12	2.41	1.63E-01	2.43E+00	1.17E+00	4.22E-01	4.08E-01	1.22E+00
95-47-6	o-xylene	1.7	2.44	11.84	1.19	2.46	1.63E-01	1.44E+00	1.48E+00	1.89E-01	2.38E-01	1.49E+00
95-63-6	1.2.4-trimethylbenzene	1.52	2.43	11.76	1.27	2.36	1.63E-01	6.43E-01	6.43E-01	1.15E-01	1.62E-01	6.43E-01
108-10-1	4-methyl-2-pentanone	1.4	2.2	10.28	2.14	2.16	5.68E-01	7.80E-01	7.17E-01	1.01E-01	1.89E-01	7.21E-01
108-83-8	2,6-dimethyl-4-heptanone	1.4	2.25	10.26	2.18	2.17	5.68E-01	9.28E+00	6.00E-01	1.91E-01	2.53E-01	4.01E-01
110-43-0	2-heptanone	1.34	2.28	10.55	2.31	2.34	5.68E-01	2.60E-01	4.80E-02	1.09E-01	2.23E-01	7.81E-02
431-03-8	2,3-butanedione (diacetyl)	0	2.77	0	0	0.12	5.68E-01	5.74E-04	6.53E-04	1.07E-01	3.71E-04	1.17E-04
589-38-8	3-hexanone	1.68	2.27	10.82	2.16	2.38	5.68E-01	5.68E-01	1.09E-01	9.79E-02	1.48E-01	1.09E-01
591-78-6	2-hexanone	1.6	2.26	10.43	2.16	2.15	5.68E-01	7.08E-01	3.59E-01	1.10E-01	2.72E-01	6.03E-01
67-64-1	acetone	5.59	1.99	7.46	2.03	2	5.68E-01	5.56E+01	4.91E+00	1.28E+00	1.95E+00	7.75E+00
78-93-3	2-butanone	4.94	2.2	10.04	2.02	2.16	5.68E-01	5.59E+00	2.35E+00	5.15E-01	1.05E+00	2.54E+00
98-86-2	acetophenone	1.48	2.92	69.92	1.45	2.78	5.68E-01	1.39E-01	1.30E-02	1.76E-02	1.80E-02	1.30E-02
105-37-3	ethyl propionate	1.67	2.48	11.12	0	2.43	2.30E-01	1.20E+00	2.52E-02	6.97E-02	8.73E-02	2.19E-02
105-54-4	ethylbutyrate	1.48	2.4	0	0	2.45	2.30E-01	1.70E-05	8.47E-05	3.46E-02	1.40E-03	5.78E-04
123-86-4	acetate de butyle	1.2	2.46	11.18	2.36	2.4	2.30E-01	1.00E-02	5.36E-02	4.34E-02	3.64E-02	5.13E-02
141-78-6	acetate d'ethyle	1.78	2.5	11.32	2.45	2.45	2.30E-01	3.31E+00	3.28E+00	3.31E-01	7.31E-01	2.91E+00
623-42-7	methylbutyrate	1.2	2.5	11.16	1.24	2.55	2.30E-01	7.69E-02	3.16E-02	7.89E-02	5.26E-02	3.32E-02
624-24-8	methyl valerate	1.17	2.46	57.97	0	2.47	2.30E-01	1.10E-02	1.05E-02	3.71E-02	4.01E-02	1.01E-02
109-99-9	tetrahydrofurane	4.42	2.34	10.26	2.2	2.35	1.64E+00	8.99E+01	2.26E+00	5.60E-01	6.73E-01	2.26E+00
1634-04-4	MTBE	1.61	1.89	41.96	1.73	1.77	1.64E+00	6.29E-01	2.50E-02	1.18E-01	9.95E-02	2.50E-02
60-29-7	diethylether	1.32	1.41	8.38	0.62	1.49	1.64E+00	7.70E-01	2.24E-01	1.88E-01	1.59E-01	1.00E-01
75-18-3	dimethylsulfide(DMS)	1.22	1.74	0.69	0	2.18	1.84E-03	5.61E-03	4.68E-03	5.04E-03	4.02E-03	4.00E-03

Il apparait que les mélanges 3 et 4 sont ceux qui contiennent le moins de molécules possédant un seuil de perception très faible (de 10^{-3} à $10^{-5} mg/m^3$), ce qui se traduit d'ailleurs par les concentrations d'odeur expérimentales les plus faibles des mélanges étudiés. Ainsi, il semble ici que plus un mélange contient de molécules possédant un seuil de perception peu élevé, plus la complexité du scénario à mettre en œuvre est élevée, et inversement.

5.3.2.2 Impact du scénario d'obtention des seuils de perception utilisé sur l'identification des molécules responsables de l'odeur

Les molécules identifiées comme responsables de l'odeur et les activités d'odeur associées sont présentées en Tableau 58 pour les mélanges 2 et 5 et illustrées par les graphes en Annexe 30 et Annexe 31.

L'approche statistique est comparée à l'approche experte, prise comme référence. Pour réaliser cette identification, la concentration des molécules utilisée pour le calcul des activités d'odeur est identique pour tous les scénarios.

Tableau 58 : Molécules identifiées comme responsables de la concentration d'odeur suivant le scénario de seuils de perception utilisé et l'approche (mélanges 2 et 5)

	Mélange 2 (CO = 31175 UO _E /m ³)											
	Scénario 1		Scénario 2		Scénario 3		Scénario 4		Scénario 5		Scénario 6	
	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}
Approche expert	dimethylsulfide	944	ethylbutyrate	141128	ethylbutyrate	28316	2-methylbutanal	458	3-methylbutanal	8046	2,3-butanedione	23564
	2-methylbutanal	113			3-methylbutanal	6367	3-methylbutanal	408	2,3-butanedione	7453	3-methylbutanal	10125
	decanal	108			2,3-butanedione	4241	dimethylsulfide	344	1-hexene	2678	ethylbutyrate	4148
	2-ethylhexanal	108					acetophenone	166	ethylbutyrate	1711	1-hexene	3587
	3-methylbutanal	106					decanal	140				
Approche statistique	dimethylsulfide	944	2,3-butanedione	4824	ethylbutyrate	28316	acetophenone	166	2,3-butanedione	7453	2,3-butanedione	23564
	3-methylbutanal	106	ethylbutyrate	141128	2,3-butanedione	4241	2-methylbutanal	458	3-methylbutanal	8046	3-methylbutanal	10125
	2-methylbutanal	113	3-methylbutanal	2967	3-methylbutanal	6367	3-methylbutanal	408	1-hexene	2678	1-hexene	3587
	decanal	108			2-methylbutanal	1411	dimethylsulfide	344	ethylbutyrate	1711		
	2-ethylhexanal	108					styrène	33	2-methylbutanal	889		
							Methyl valérate	66				
						decanal	140					

	Mélange 5 (CO = 17653 UO _E /m ³)											
	Scénario 1		Scénario 2		Scénario 3		Scénario 4		Scénario 5		Scénario 6	
	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}	Molécule	OAV _{exp}
Approche expert	dimethylsulfide	1184	ethylbutyrate	143953	ethylbutyrate	28882	dimethylsulfide	432	3-methylbutanal	1849	octanal	4503
	decanal	109					acetophenone	158	ethylbutyrate	1745	ethylbutyrate	4231
	2-ethylhexanal	107					decanal	141	1-hexene	1259	3-methylbutanal	2327
	octanal	107					2-methylbutanal	120	dimethylsulfide	542	1-hexene	1686
							octanal	94			2,3-butanedione	1060
Approche statistique							3-methylbutanal	94				
	dimethylsulfide	1184	ethylbutyrate	143953	ethylbutyrate	28882	acetophenone	158	2,3-butanedione	335	2,3-butanedione	1060
	2-ethylhexanal	107			3-methylbutanal	1463	2-methylbutanal	120	3-methylbutanal	1849	3-methylbutanal	2327
					2-methylbutanal	369	3-methylbutanal	94	acetophenone	155	2-methylbutanal	369
					dimethylsulfide	465	dimethylsulfide	432	1-hexene	1259	methyl valerate	246
							methyl valerate	67	ethylbutyrate	1745	acetophenone	215
						styrène	33	2-methylbutanal	232			

Concernant l'approche de l'expert, on peut constater que le scénario 5 est celui qui permet d'approcher au mieux les résultats du scénario 6 pris comme référence.

- Pour le mélange 2, les molécules identifiées sont identiques mais l'ordre diffère.
- Pour le mélange 5, l'identification présente des discordances plus importantes (absence de l'octanal et de la 2,3-butanedione, présence du diméthylsulfide). C'est néanmoins le scénario qui permet de se rapprocher au mieux des résultats obtenus avec le scénario 6.

L'application de l'approche statistique fait elle aussi ressortir une concordance globale des scénarios 5 et 6 même si certaines discordances peuvent être constatées. Pour le mélange 5, deux molécules ressortent, bien que non identifiées par l'expert : le méthyl valérate (scénario 6) et l'acétophénone (scénario 5 et 6).

Comme mentionné précédemment, l'approche statistique fait ressortir ici les molécules expliquant le mieux la variation de la concentration d'odeur de l'ensemble des données utilisées, et non spécifiquement les molécules responsables de l'odeur du mélange étudié.

5.3.3 Bilan sur le scénario d'obtention des seuils de perception

Cette étude permet de montrer que parmi l'ensemble des scénarios envisagés pour obtenir les seuils de perception, le scénario 5, basé sur une modélisation locale, est celui qui permet d'approcher au mieux les valeurs de seuils utilisées par l'expert (scénario 6). Il permet donc logiquement le plus fréquemment d'approcher la concentration d'odeur expérimentale des mélanges ainsi que l'identification de l'expert.

La limite de ces résultats, évoquée en partie 5.3.2.1, réside dans le fait que la majorité des molécules étudiées ici sont utilisées dans la construction du modèle et du scénario 5.

Au final, les identifications des molécules responsables de l'odeur des échantillons industriels analysés dans la suite de ce travail (chapitre 6) seront donc réalisées en parallèle avec les scénarios 5 et 6.

5.4 Évaluation de l'impact de la précision de l'analyse olfactométrique

5.4.1 Méthodologie

Comme présenté au chapitre 1, la concentration d'odeur d'un mélange gazeux est mesurée conformément à la norme EN 13725 (AFNOR, 2003). Cette norme stipule qu'un jury de 4 jurés certifiés au n-butanol permet de réaliser une mesure représentative de la concentration d'odeur. Pour rappel, un juré est dit certifié au n-butanol si ce juré perçoit l'odeur du n-butanol à une concentration comprise entre 0,02 et 0,08 $\mu\text{mol/mol}$ (AFNOR, 2003).

Dans cette partie de l'étude, il s'agit d'observer l'impact du nombre de jurés et de leur certification sur la mesure de la concentration d'odeur et par voie de conséquence sur l'identification des molécules responsables de l'odeur.

Deux mélanges synthétiques ont été choisis pour réaliser cette partie de l'étude : les mélanges 2 et 5. Ils ont été choisis en raison du fait qu'il s'agit des mélanges présentant les concentrations d'odeur les plus élevées, ce qui facilite la réalisation de plusieurs analyses olfactométriques sur un seul et même échantillon. Pour réaliser cette étude, les concentrations expérimentales et les seuils de perception du scénario 6 sont utilisés.

Trois séances d'olfactométrie ont été réalisées, chacune avec 6 jurés:

- Une première avec 6 jurés certifiés au n-butanol (Jury 1);
- Une seconde avec 6 jurés présentant une hyperosmie au n-butanol (Jury 3, non certifié) ;
- Une troisième avec 6 jurés présentant une hyposmie au n-butanol (Jury 4, non certifié).

La mesure a été quant à elle systématiquement réalisée selon la norme EN 13725.

Toutes les mesures de concentration d'odeur réalisées dans cette partie de l'étude l'ont été en plus de celles déjà précédemment réalisées sur les mélanges 2 et 5 avec un jury certifié (Jury 2).

L'impact du nombre de jurés et de leur certification ou non sur le résultat de la mesure de la concentration d'odeur est examiné dans un premier temps. L'impact du protocole d'obtention de la concentration d'odeur sur l'identification des molécules responsables de l'odeur est ensuite analysé à l'aide de l'outil statistique.

5.4.2 Résultats

Les différentes analyses réalisées sur les mélanges 5 et 8, tant spécifiquement pour cette partie de l'étude que précédemment, permettent d'évaluer dans un premier temps l'impact du protocole d'obtention de la concentration d'odeur (jurés certifiés ou non, nombre de jurés) sur la valeur obtenue. Cet impact est mesuré par le taux de variation de la concentration d'odeur $\left(\frac{CO_{th}-CO_{exp}}{CO_{exp}}\right) \times 100$ en prenant comme référence les valeurs de concentration d'odeur obtenues avec le jury 1.

5.4.2.1 Impact du nombre et de la nature des jurés sur la concentration d'odeur

5.4.2.1.1 Jurys certifiés

Les valeurs de concentrations d'odeur obtenues pour les différentes configurations de jurys certifiés sont résumées dans le Tableau 59.

Tableau 59 : Comparaison des valeurs de concentration d'odeur (UOE/m³) obtenues avec les jurys certifiés (6 et 12 jurés)

	6 jurés certifiés			12 jurés certifiés	
	Jury 1 (Référence)	Jury 2	Taux de variation (%)	Jury 1 + Jury 2	Taux de variation (%)
Mélange 2	28301	31175	10	28694	1
Mélange 5	14316	17653	23	14751	3

L'examen des résultats présentés dans le Tableau 59 fait ressortir que le taux de variation de la concentration d'odeur obtenue avec un jury de 6 personnes varie de 10 à 23 % selon le mélange considéré. Il apparaît aussi que l'utilisation d'un jury certifié de 12 personnes (Jury 1 + Jury 2) se traduit par une nette amélioration du taux de variation de la concentration d'odeur (1 à 3%).

5.4.2.1.2 Jurys non certifiés

Les valeurs de concentrations d'odeur obtenues pour les différentes configurations de jurys non certifiés sont résumées dans les Tableau 60 et Tableau 61.

Tableau 60 : Comparaison des valeurs de concentration d'odeur (UOE/m³) obtenues avec les jurys non certifiés (6 et 12 jurés)

	6 jurés certifiés		6 jurés non certifiés				12 jurés non certifiés	
	Jury 1	Jury 3	Taux de variation (%)	Jury 4	Taux de variation (%)	Jury 3 + Jury 4	Taux de variation (%)	
Mélange 2	28301	30572	8	20548	-27	25064	-11	
Mélange 5	14316	15501	8	9964	-30	12680	-11	

On note pour le jury 3 un taux de variation de la concentration d'odeur logiquement positif de 8 % (jury hyperosmique), indépendamment du mélange considéré. On peut observer ici que ce taux est inférieur à celui observé entre les deux jurys certifiés (Tableau 59). Le taux de variation est aussi logiquement négatif et se situe entre 27 et 30% pour le jury 4 (hyposmique).

Enfin, l'utilisation d'un jury de 12 personnes non certifiées (Jury 3 + Jury 4) se traduit par un taux de variation de 11 % par rapport aux valeurs de référence, logiquement négatif, l'hyposmie du jury 4 ayant plus de poids sur la valeur de la concentration d'odeur que l'hyperosmie du jury 3.

Les valeurs de concentrations d'odeur obtenues avec un jury non certifié de 24 personnes, constitué de 12 jurés certifiés et de 12 jurés non certifiés (6 hyposmiques et 6 hyperosmiques) sont présentées dans le Tableau 61.

Tableau 61 : Comparaison des valeurs de concentration d'odeur (UOE/m³) obtenues avec les jurys non certifiés (24 jurés)

	6 jurés certifiés	24 jurés (Jury non certifié)	
	Jury 1 (Référence)	Jury 1 + Jury 2 + Jury 3 + Jury 4	Taux de variation (%)
Mélange 2	28301	26764	-5
Mélange 5	14316	14056	-2

L'utilisation d'un jury de 24 personnes (Jury 1 + Jury 2 + Jury 3 + Jury 4) se traduit par un taux de variation de 2 à 5 % par rapport aux valeurs de référence, encore une fois logiquement négatif, au égard du poids plus important de l'hyposmie du jury 4 sur la valeur de la concentration d'odeur.

5.4.2.1.3 Bilan de l'impact du nombre et de la nature des jurés sur la précision de la mesure de la concentration d'odeur

Les valeurs de taux de variation de la concentration d'odeur obtenus à l'aide des différents types de jurys sont résumées dans le Tableau 62.

Tableau 62 : Comparaison des taux de variation des concentrations d'odeur (UOE/m³) obtenues avec les différents types de jurys

	6 Jurés			12 Jurés		24 Jurés
	certifié	non	non	certifiés	non certifiés	non certifiés
	Jury 2	Jury 3	Jury 4	Jury 1 + Jury	Jury 3 + Jury	Jury 1 + Jury 2 + Jury 3 + Jury
Mélange 2	10	8	-27	1	-11	-5
Mélange 5	23	8	-30	3	-11	-2
Taux de variation moyen	17	8	-29	2	-11	-4

Ces résultats permettent de dégager les conclusions suivantes :

- Pour les jurys constitués de 6 jurés, les meilleurs résultats sont obtenus avec le jury non certifié hyperosmique (+8% en moyenne). Les plus éloignés des valeurs prises comme référence sont celles obtenues avec le jury hyposmique (-29% en moyenne).
- Le doublement du nombre de jurés se traduit par une très nette diminution du taux de variation pour le jury certifié (2% en moyenne) ainsi que pour le jury non certifié. Le taux de variation obtenu (-11%) est du même ordre de grandeur que celui obtenu pour un jury certifié de 6 personnes (17%).
- Enfin, l'utilisation d'un jury de 24 personnes constitué à parts égales d'éléments certifiés et non certifiés conduit à un résultat du même niveau de qualité (-4%) que celui obtenu avec 12 jurés certifiés (2%).

5.4.2.2 Impact de la précision de la concentration d'odeur sur l'identification des molécules responsables de l'odeur

Comme expliqué au chapitre 4, pour l'expert, la concentration d'odeur mesurée sert de référence. Il compare la somme des activités d'odeur des molécules identifiées (concentration d'odeur « théorique ») à cette valeur pour s'assurer de la pertinence de son identification.

La précision sur la mesure olfactométrique n'influe pas directement sur son identification et sa hiérarchisation des molécules responsables de l'odeur (il se base uniquement sur les activités d'odeur).

Pour rappel, les molécules que son analyse fait ressortir par ordre d'importance dans les mélanges 2 et 5 sont respectivement (Chapitre 4, tableau 11) :

- 2,3-butanedione, 3-methylbutanal, ethylbutyrate, 1-hexene
- Octanal, ethylbutyrate, 3-methylbutanal, 1-hexene, 2,3-butanedione

Il en va différemment pour l'analyse statistique. En effet, cette dernière utilise en plus les résultats des mesures olfactométriques dans le processus d'identification.

Les résultats précédents ont montré que le taux de variation de la concentration d'odeur peut varier de 2 à 30% suivant le type de jury utilisé :

- +2% pour un jury de 12 personnes certifiées
- -11% pour un jury de 12 personnes non certifiées
- -29% pour un jury de 6 personnes non certifiées (hyposmiques)

Ce sont donc ces trois situations qui vont être utilisées afin d'observer l'impact de la variation de la concentration d'odeur sur l'identification réalisée à l'aide de l'outil statistique. Le Tableau 63 ci-dessous présente les molécules identifiées par ordre de VIP décroissante (Annexe 32) pour les deux mélanges et les trois niveaux de précision. L'identification obtenue sur la base des résultats du jury 1 est prise comme référence.

Tableau 63 : Approche statistique : Molécules identifiées comme responsables de la concentration d'odeur (mélanges 2 et 5)

		Taux de variation par rapport à la valeur de référence		
	Référence (Jury 1)	2%	-11%	-29%
Mélange 2	2,3-butanedione	2,3-butanedione	2,3-butanedione	2,3-butanedione
	3-methylbutanal	3-methylbutanal	3-methylbutanal	3-methylbutanal
	1-hexene	2-methylbutanal	2-methylbutanal	2-methylbutanal
		1-hexene	1-hexene	
Mélange 5	2,3-butanedione	2,3-butanedione	2,3-butanedione	2,3-butanedione
	3-methylbutanal	2-methylbutanal	2-methylbutanal	2-methylbutanal
	2-methylbutanal	3-methylbutanal	3-methylbutanal	3-methylbutanal
	methyl valerate	methyl valerate	methyl valerate	methyl valerate
	acetophenone	acetophenone	acetophenone	acetophenone
		dimethylsulfide	dimethylsulfide	dimethylsulfide
		octanal		

L'examen des résultats fait ressortir que dans les conditions de notre étude, pour le mélange 5, aucune différence significative n'apparaît dans l'identification globale quel que soit le taux de variation. Même si des molécules supplémentaires apparaissent (dimethylsulfide et octanal), toutes les molécules de la référence sont présentes.

Le même constat peut être fait dans le cas du mélange 2 pour les taux de variation de +2 et -11%. Pour un taux de variation de -29%, le 1-hexène, présent dans l'identification de référence, disparaît.

Pour les deux mélanges, une modification de la hiérarchie des molécules identifiées est notée quel que soit le taux de variation.

Enfin, sur un plan général, la comparaison des identifications issues de l'approche statistique avec celles de l'expert fait ressortir à nouveau les discordances soulignées au chapitre 4.

5.5 Conclusion intermédiaire sur l'impact de la qualité des données utilisées sur la détermination des molécules responsables de l'odeur

À l'issue de cette partie de l'étude, des conclusions peuvent être dégagées, pour chacune des deux approches, quant à l'impact du niveau de précision des données utilisées sur la détermination des molécules responsables de l'odeur d'un mélange gazeux.

Approche expert :

L'amélioration de la quantification fait que les activités d'odeur calculées sont plus représentatives du poids réel des molécules dans la concentration d'odeur du mélange. Pour les deux mélanges étudiés, seule la quantification absolue permet d'obtenir une identification et une hiérarchisation conformes à la référence.

Parmi l'ensemble des scénarios envisagés pour obtenir les seuils de perception, le scénario 5 (modélisation locale) ressort comme étant celui qui permet d'approcher au mieux les résultats de l'expert (scénario 6).

Enfin, la précision sur la mesure olfactométrique n'influe pas sur l'identification et la hiérarchisation des molécules responsables de l'odeur par l'expert. La concentration d'odeur mesurée lui sert uniquement de référence pour s'assurer de la pertinence de son identification.

Approche statistique :

La PLS fait ressortir les molécules expliquant au mieux la variation de la concentration d'odeur de l'ensemble des données utilisées, et non spécifiquement celles responsables de l'odeur du mélange étudié. Une amélioration de la précision des données d'entrée (quantifications physicochimiques et olfactométriques, seuils de perception) devrait conduire à une meilleure qualité du résultat. Cependant, cette amélioration s'avère difficilement discernable dans le cadre de notre étude.

Enfin, il convient de souligner que ces résultats ont été obtenus pour des mélanges étalons constitués d'une proportion très importante (>75%) de molécules dont les seuils de perception avaient été mesurés dans le cadre de cette étude. Une telle proportion sera clairement inatteignable lors d'une analyse courante. Comme mentionné au chapitre 3, le nombre de seuils de perception disponibles dans la

littérature est très faible en regard du nombre de molécules susceptibles d'être identifiées. Nos résultats ayant montré que la modélisation locale (scénario 5) permettait d'approcher au mieux les résultats de l'expert (scénario 6), il apparaîtrait donc pertinent d'envisager un nouveau scénario les combinant. Le nouveau scénario ainsi élaboré consisterait, pour chaque molécule identifiée, à utiliser prioritairement le seuil de perception mesuré au laboratoire, puis la moyenne géométrique des seuils de perception disponibles dans la base de données nettoyée, et enfin la valeur issue de la modélisation locale si la molécule n'est pas renseignée dans la base. Cette approche présenterait l'avantage de ne pas avoir à classer la molécule dans une famille chimique donnée pour lui attribuer un seuil moyen, chose en outre parfois délicat à réaliser, en particulier dans le cas des molécules polyfonctionnelles.

CHAPITRE 6 : EXEMPLE D'APPLICATION DE LA DEMARCHE

6.1 Contexte général et objectif

L'objectif de cette partie de l'étude est d'appliquer la méthode développée dans les chapitres précédents sur un cas réel afin d'évaluer l'outil avec des données plus incertaines. Jusqu'ici la démarche a en effet été développée et appliquée sur des mélanges gazeux synthétiques, ce qui a permis de rechercher le niveau de précision des données nécessaire pour obtenir une identification satisfaisante des molécules responsables de l'odeur.

Au chapitre 5, il a été montré que le nombre de jurés utilisés lors des analyses olfactométriques ou leur certification selon la norme EN 13725 ne modifient pas significativement l'identification des molécules responsables de l'odeur. Cette identification est par contre influencée par la précision de la quantification des molécules présentes dans le gaz ainsi que celle des seuils de perception. Le scénario qui a fourni les résultats les plus proches de ceux de l'expert (scénario 6), pris comme référence, est le scénario de modélisation locale (scénario 5). Toutefois, la construction de ce modèle se base principalement sur les valeurs de seuils utilisées dans le scénario expert (scénario 6). Il est donc important de valider ce modèle sur des molécules différentes de celles utilisées pour la construction.

Dans ce chapitre, seuls les résultats de l'approche « automatisée » de l'expert seront présentés. En effet, comme l'ont montré les résultats présentés dans les chapitres 4 et 5, l'approche statistique n'apporte pas une amélioration significative de l'identification des molécules responsables de l'odeur.

La capacité de l'outil développé à identifier les molécules responsables de l'odeur est évaluée sur la base de l'analyse de quatre matériaux. Cette évaluation est réalisée avec les scénarios 5 et 6, à partir des données quantifiées le plus précisément possible. Elle est complétée par la mise en œuvre d'un scénario 7 combinant les deux précédents. Comme exposé au chapitre 5, il consiste, pour chaque molécule identifiée, à utiliser prioritairement le seuil de perception mesuré au laboratoire, puis la moyenne géométrique des seuils de perception disponibles dans la base de données nettoyée, et enfin la valeur issue de la modélisation locale si la molécule n'est pas renseignée dans la base.

Dans un premier temps, la méthodologie utilisée pour l'étude est présentée, puis les résultats et les conclusions associées sont détaillés.

6.2 Matériels et méthodes

6.2.1 Matériaux analysés

L'étude a été réalisée sur quatre échantillons de matériaux caoutchouc (deux granulométries différentes / avant et après vulcanisation). Elle a été conduite de manière à renseigner la composition de la matrice gazeuse de la manière la plus complète et pertinente possible, et à évaluer la concentration d'odeur.

Tableau 64 : Matériaux analysés

Caoutchouc cru
Caoutchouc cru broyé
Caoutchouc vulcanisé
Caoutchouc vulcanisé broyé

6.2.2 Protocole de l'étude

L'étude a été réalisée dans une pièce à température régulée (20°C). Le protocole expérimental utilisé a été le suivant :

1. $t=0$: Conditionnement des échantillons dans des sacs Nalophan®.
2. Vérification de l'atteinte de l'équilibre d'émission des matériaux par des mesures périodiques des COV émis (détecteur à photoionisation portable - ppb RAE Systems).
3. Réalisation des analyses physicochimiques et olfactométriques.

6.2.2.1 Protocole de conditionnement en sacs Nalophan®

Chaque échantillon de matériau est introduit dans un sac en Nalophan® qui est ensuite rempli avec 40L d'air de référence (air pur et sec) et placé dans un pièce à température régulée ($T= 20 \pm 2^\circ\text{C}$) (Figure 67).



Figure 67 : Sacs en Nalophan®

L'étape suivante consiste à enregistrer régulièrement la concentration en COV dans le sac jusqu'à l'atteinte de l'équilibre d'émission (équilibre thermodynamique). Ces mesures ont été réalisées avec un détecteur à photoionisation ppb RAE Systems.

6.2.2.2 Analyses olfactométriques

Une fois l'équilibre d'émission atteint, les analyses olfactométriques sont réalisées. Elles ont permis d'évaluer la persistance (concentration d'odeur – NF EN 13725) de l'odeur émise par les matériaux. Pour mesurer la concentration de l'odeur, le gaz odorant à analyser (présent à l'intérieur des sacs) a été présenté à différentes concentrations à un jury de six personnes sous la forme de dilutions successives.

Cette analyse sensorielle a été effectuée avec à un olfactomètre à dilution dynamique multiposte Odile® (Figure 68) et respecte la norme NF EN 13725.



Figure 68 : Olfactomètre à Dilution Dynamique Odile®

Pour chaque dilution, les jurés indiquent s'ils ont perçu l'odeur ou non. Ainsi pour chaque juré puis pour le jury, le seuil de perception (concentration d'odeur) des odeurs peut être déterminé.

6.2.2.3 Analyses physico-chimiques

La méthode utilisée pour identifier les composés organiques volatils est basée sur l'utilisation d'un tube d'adsorbant permettant de réaliser une préconcentration (Tubes d'échantillonnage ATD en acier inoxydable préremplis). La plage de volatilité des composés dans les tubes couvre approximativement la gamme de n-C3 à n-C12. La préconcentration est réalisée en faisant passer un volume connu d'air sur un tube Air Toxics® à un débit régulé de 50 mL/mn (Pompe GILIAN GILAIRPlus).

La thermodésorption est réalisée à l'aide d'un thermodésorbeur Turbomatrix 350 (Perkin Elmer). La chaîne analytique utilisée y associe un chromatographe en phase gazeuse et un spectromètre de masse (GC CLARUS 680 / SQ8T, Perkin Elmer).

Le tube est désorbé à 300°C pendant 15 min sous un flux d'hélium 6,0 à 40 mL/min. Les composés désorbés sont concentrés sur un piège froid maintenu à -30°C. À la fin de cette désorption primaire, ce piège froid est flashé thermiquement (40°C/s jusqu'à 300°C) pour l'injection dans le chromatographe.

Les composés identifiés font l'objet :

1. D'une quantification absolue lorsque nous disposons d'un étalonnage pour la molécule concernée (résultat exprimé en mg/m³).
2. D'une semi-quantification en mg/m³ équivalent « famille chimique » lorsque nous ne disposons pas d'un étalonnage pour la molécule concernée mais que nous disposons d'étalonnages pour d'autres molécules de la même famille chimique. Dans ce cas, c'est le coefficient directeur moyen des différentes droites d'étalonnage dont nous disposons qui est employé.
3. D'une semi-quantification en mg/m³ équivalent toluène lorsque nous ne disposons d'aucun étalonnage pour la famille chimique à laquelle appartient la molécule concernée.

6.3 Résultats

6.3.1 Analyses physico-chimiques

En Tableau 65 est présenté le nombre de molécules identifiées dans les émissions de chaque matériau.

Tableau 65 : Nombre de molécules présentes dans les émissions des matériaux

Échantillons	Nombre de molécules présentes
Caoutchouc cru	51
Caoutchouc cru broyé	38
Caoutchouc vulcanisé	36
Caoutchouc vulcanisé broyé	35

Les concentrations sont présentées en Tableau 66. Pour des raisons de confidentialité, les molécules présentes dans les échantillons sont anonymisées.

Tableau 66 : Concentrations des molécules présentes dans les différents échantillons

Molécule	Famille	Caoutchouc cru	Caoutchouc cru broyé	Caoutchouc vulcanisé	Caoutchouc vulcanisé broyé	Unités
1	Acides	0,335	0,411	0,0605	0,0807	mg éq. Toluène/m ³
2	Alcanes	2,73	0,82	0,934	0,716	mg éq. Alcanes/m ³
3	Alcanes	0,0341	0,0436	0,105	0,0586	mg/m ³
4	Alcanes	nd	nd	0,0716	0,0714	mg/m ³
5	Alcanes	0,133	nd	0,0605	0,0285	mg éq. Alcanes/m ³
6	Alcanes	nd	nd	0,294	0,248	mg éq. Alcanes/m ³
7	Alcanes	0,157	0,109	0,128	0,139	mg/m ³
8	Alcanes	0,189	0,0647	0,0793	0,0641	mg éq. Alcanes/m ³
9	Alcanes	nd	nd	0,0205	0,0175	mg éq. Alcanes/m ³
10	Alcanes	0,442	0,417	0,436	nd	mg éq. Alcanes/m ³
11	Alcanes	0,432	0,147	nd	nd	mg éq. Alcanes/m ³
12	Alcanes	0,507	0,367	nd	nd	mg éq. Alcanes/m ³
13	Alcanes	0,0435	0,0273	nd	nd	mg éq. Alcanes/m ³
14	Alcanes	0,0041	0,0128	nd	nd	mg éq. Alcanes/m ³
15	Alcanes	0,0604	0,0428	nd	nd	mg éq. Alcanes/m ³
15	Alcanes	0,0604	0,0428	nd	nd	mg éq. Alcanes/m ³
16	Alcanes	0,0784	0,0245	nd	nd	mg éq. Alcanes/m ³
17	Alcanes	nd	0,0204	nd	nd	mg éq. Alcanes/m ³
18	Alcanes	0,00778	nd	nd	nd	mg éq. Alcanes/m ³
19	Alcanes	0,0148	nd	nd	nd	mg éq. Alcanes/m ³
20	Alcanes	0,00635	nd	nd	nd	mg éq. Alcanes/m ³
21	Alcanes	0,006	nd	nd	nd	mg éq. Alcanes/m ³
22	Alcanes	0,0328	nd	nd	nd	mg éq. Alcanes/m ³
23	Alcanes	0,0431	nd	nd	nd	mg éq. Alcanes/m ³
24	Alcanes	0,00932	nd	nd	nd	mg/m ³
25	Alcanes	0,0217	nd	nd	nd	mg éq. Alcanes/m ³
26	Alcènes	nd	nd	0,0205	0,0143	mg éq. Alcènes/m ³
27	Alcènes	0,00419	0,016	nd	0,0158	mg éq. Alcènes/m ³
28	Alcènes	0,0143	0,0375	nd	nd	mg éq. Alcènes/m ³
29	Alcènes	0,0347	nd	nd	nd	mg éq. Alcènes/m ³
30	Alcènes	nd	nd	nd	0,0128	mg éq. Alcènes/m ³
31	Alcools	nd	nd	0,242	nd	mg/m ³
32	Aromatiques	0,256	0,283	39,3	37	mg éq. Alcools/m ³
33	Alcools	53,9	110	65,1	46,5	mg/m ³
34	Alcools	0,0475	0,0756	3,2	1,41	mg éq. Alcools/m ³
35	Aldéhydes	0,0121	0,0279	0,0298	0,0323	mg éq. Aldéhyde/m ³
36	Aldéhydes	nd	nd	0,0906	0,0422	mg éq. Aldéhyde/m ³
37	Aldéhydes	nd	nd	0,394	nd	mg éq. Aldéhyde/m ³
38	Aldéhydes	nd	0,153	0,0602	0,0564	mg éq. Aldéhyde/m ³
39	Aldéhydes	1,44	2,96	2	3,13	mg éq. Aldéhyde/m ³
40	Aldéhydes	0,0133	0,0297	nd	nd	mg éq. Aldéhyde/m ³
41	Aromatiques	0,00046	0,0312	0,0364	0,0355	mg/m ³
42	Aromatiques	0,15	0,169	0,118	0,0932	mg/m ³
43	Aromatiques	nd	nd	0,269	0,171	mg éq. Aromatiques/m ³
44	Aromatiques	nd	nd	1,62	1,92	mg éq. Ethers/m ³
45	Aromatiques	0,0107	nd	0,169	0,0996	mg éq. Aromatiques/m ³
46	Aromatiques	0,191	0,347	82,4	69,3	mg éq. Aromatiques/m ³
47	Aromatiques	0,00508	0,00877	nd	nd	mg/m ³
48	Aromatiques	0,00513	0,0201	nd	nd	mg/m ³
49	Aromatiques	0,000609	0,02	nd	nd	mg/m ³
50	Aromatiques	nd	0,0143	nd	nd	mg éq. Aromatiques/m ³
51	Azotés	0,137	0,125	nd	nd	mg éq. Toluène/m ³
52	Cétones	0,357	0,835	1,69	1,21	mg/m ³
53	Cétones	0,164	0,272	48,8	50,1	mg/m ³
54	Cétones	0,0375	0,0361	nd	nd	mg éq. Cétones/m ³
55	Cétones	0,00558	nd	nd	nd	mg/m ³
56	Chlorés	0,51	1,21	0,894	0,214	mg éq. Toluène/m ³
57	Chlorés	0,0167	nd	nd	nd	mg éq. Toluène/m ³
58	Esters	0,136	0,187	1,87	0,0861	mg/m ³
59	Esters	nd	nd	0,711	0,453	mg éq. Esters/m ³
60	Esters	0,00442	nd	nd	nd	mg éq. Esters/m ³
61	Ethers	0,00226	nd	0,116	0,0352	mg éq. Ethers/m ³
62	Ethers	0,0826	0,129	0,107	0,131	mg éq. Ethers/m ³
63	Ethers	0,144	nd	0,0124	0,0186	mg/m ³
64	Soufrés	nd	nd	0,204	0,217	mg/m ³
65	Soufrés	0,00296	0,0663	2,16	3,17	mg éq. Soufrés/m ³

Parmi les soixante-cinq molécules analysées, 17 (26%) ont fait l'objet d'une quantification absolue, 45 (69%) ont été quantifiées en équivalent-famille, et 4 (6%) ont été quantifiées en équivalent-toluène.

6.3.1 Analyses olfactométriques

Les concentrations d'odeur mesurées pour chacun des échantillons sont présentées dans le Tableau 67.

Tableau 67 : Concentration d'odeur des échantillons

Echantillons	CO (UO_E/m^3)
Caoutchouc cru	497
Caoutchouc cru broyé	1005
Caoutchouc vulcanisé	3216
Caoutchouc vulcanisé broyé	2866

Les quatre matériaux étudiés présentent des concentrations d'odeur globalement faibles, comprises entre 497 et 3216 UO_E/m^3 . Les matériaux vulcanisés présentent des niveaux d'odeur significativement supérieurs aux autres.

6.3.2 Seuils de perception

Les valeurs de seuils de perception obtenues par application des scénarios 5, 6 et 7 sont synthétisées en Tableau 68.

Tableau 68 : Valeurs des seuils de perception obtenues suivant les scénarios

Molécule	Famille	Expert (Scénario 6) (mg/m^3)	Source	Modèle (Scénario 5) (mg/m^3)	Scénario 7 (mg/m^3)	Source
1	Acides	1,90000	BDDfam	0,00743	0,00743	Modèle
2	Alcanes	0,60000	BDD	2,38864	0,60000	BDDn
3	Alcanes	12,40253	Labo	3,11495	12,40253	Labo
4	Alcanes	12,21900	BDD	3,19053	9,88872	BDDn
5	Alcanes	13,57932	BDDfam	18,23361	18,23361	Modèle
6	Alcanes	13,57932	BDDfam	2,40495	2,40495	Modèle
7	Alcanes	57,14031	BDD	3,27863	2,70000	BDDn
8	Alcanes	3,40000	BDD	2,50622	3,40000	BDDn
9	Alcanes	1,70000	BDD	2,69468	1,70000	BDDn
10	Alcanes	13,57932	BDDfam	2,55811	2,55811	Modèle
11	Alcanes	29,56849	BDD	29,55647	37,20939	BDDn
12	Alcanes	13,57932	BDDfam	2,55811	2,55811	Modèle
13	Alcanes	13,57932	BDDfam	11,73760	11,73760	Modèle
14	Alcanes	13,57932	BDDfam	8,16374	8,16374	Modèle
15	Alcanes	13,57932	BDDfam	11,73760	11,73760	Modèle
16	Alcanes	18,00000	BDD	14,29695	18,00000	BDDn
17	Alcanes	8,00000	BDD	5,04450	8,00000	BDDn
18	Alcanes	13,57932	BDDfam	1,88862	1,88862	Modèle
19	Alcanes	13,57932	BDDfam	1,65732	1,65732	Modèle
20	Alcanes	13,57932	BDDfam	3,88815	3,88815	Modèle
21	Alcanes	13,57932	BDDfam	7,03710	7,03710	Modèle
22	Alcanes	13,57932	BDDfam	31,91982	31,91982	Modèle
23	Alcanes	13,57932	BDDfam	2,17912	2,17912	Modèle
24	Alcanes	31,00000	BDD	3,05725	31,00000	BDDn
25	Alcanes	5,80000	BDD	4,82160	5,80000	BDDn
26	Alcènes	0,00718	Labo	0,00904	0,00718	Labo
27	Alcènes	0,50992	BDDfam	0,13529	0,13529	Modèle
28	Alcènes	0,50992	BDDfam	0,35551	0,35551	Modèle
29	Alcènes	0,50992	BDDfam	0,43254	0,43254	Modèle
30	Alcènes	0,50992	BDDfam	0,00245	0,00245	Modèle
31	Alcools	0,38087	Labo	0,24031	0,38087	Labo
32	Aromatiques	1,15027	BDDfam	0,03467	0,03467	Modèle

33	Alcools	0,09358	Labo	0,11781	0,09358	Labo
34	Alcools	32,62231	BDD	0,70553	12,05636	BDDn
35	Aldéhydes	0,03564	BDD	0,00614	0,00973	BDDn
36	Aldéhydes	0,01806	BDD	0,00490	0,00201	BDDn
37	Aldéhydes	0,00050	Labo	0,01047	0,00050	Labo
38	Aldéhydes	0,00325	Labo	0,00973	0,00325	Labo
39	Aldéhydes	0,03002	BDD	0,12511	0,07881	BDDn
40	Aldéhydes	0,04649	BDD	0,02245	0,02012	BDDn
41	Aromatiques	0,05920	Labo	0,07574	0,05920	Labo
42	Aromatiques	1,21658	Labo	0,40767	1,21658	Labo
43	Aromatiques	17,05479	BDD	2,21234	8,80920	BDDn
44	Aromatiques	0,16353	BDDfam	3,62150	3,62150	Modèle
45	Aromatiques	0,15323	BDD	0,10649	0,04053	BDDn
46	Aromatiques	2,19618	BDD	0,04511	0,25000	BDDn
47	Aromatiques	0,31642	BDD	0,16576	0,14124	BDDn
48	Aromatiques	0,57906	BDD	0,22755	0,36075	BDDn
49	Aromatiques	0,50822	BDD	0,22633	0,48611	BDDn
50	Aromatiques	0,16353	BDDfam	0,05126	0,05126	Modèle
51	Azotés	1946,22701	BDD	19,91827	22,00000	BDDn
52	Cétones	7,75479	Labo	1,94769	7,75479	Labo
53	Cétones	0,01295	Labo	0,01799	0,01295	Labo
54	Cétones	9,20488	BDD	0,07950	1,30000	BDDn
55	Cétones	2,54393	Labo	1,04972	2,54393	Labo
56	Chlorés	11,36047	BDD	7,33288	11,29472	BDDn
57	Chlorés	580,26771	BDD	1,15618	639,84344	BDDn
58	Esters	2,90913	Labo	0,73074	2,90913	Labo
59	Esters	0,21619	BDDfam	0,01664	0,01664	Modèle
60	Esters	90,73796	BDD	4,08438	8,10000	BDDn
61	Ethers	1,18579	BDDfam	3,01763	3,01763	Modèle
62	Ethers	1,18579	BDDfam	5,08149	5,08149	Modèle
63	Ethers	0,10002	Labo	0,15852	0,10002	Labo
64	Soufrés	0,00567	BDD	0,00409	0,00569	BDDn
65	Soufrés	0,11497	BDD	0,40185	0,10100	BDDn
BDD : Base De Données ; BDDn : Base De Données Nettoyée ; BDDfam : Famille chimique de la molécule ; Labo : Mesure Laboratoire ; Modèle : Modélisation affinée						

Parmi les soixante-cinq molécules analysées, 13 (20%) ont un seuil de perception qui a été mesuré au laboratoire, 28 (43%) ont un seuil renseigné dans la littérature, et enfin, 24 (37%) ne sont pas renseignées.

6.4 Discussion

Les résultats de l'identification des molécules responsables de l'odeur des 4 matériaux à l'aide des 3 scénarios d'obtention des seuils sont présentés dans le Tableau 69.

L'approche de l'expert (scénario 6) nous servira de référence pour la validation de la méthode.

Tableau 69 : Identification des molécules responsables de l'odeur : résultats de l'approche expert automatisée

Echantillons	Scénario 5		Scénario 6			Scénario 7		
	Code molécule	AO	Code molécule	Source SP	AO	Code molécule	Source SP	AO
Caoutchouc cru	33	458	33	Labo	576	33	Labo	576
	1	45	39	BDD	48	1	Modèle	45
Caoutchouc cru broyé	33	934	33	Labo	1175	33	Labo	1175
	1	55	39	BDD	99			
Caoutchouc vulcanisé	53	2712	53	Labo	3768	53	Labo	3768
	46	1827	37	Labo	788	32	Modèle	1133
	32	1133	33	Labo	696	37	Labo	788
	33	553				33	Labo	696
Caoutchouc vulcanisé broyé	53	2745	53	Labo	3869	53	Labo	3869
	46	1536	33	Labo	497	32	Modèle	1067
	32	1067				33	Labo	497
	33	395				46	BDDn	277

Le Tableau 69 permet d'observer que, globalement, les résultats obtenus à l'aide du scénario 5 et du scénario 6 font ressortir la même molécule ayant l'activité d'odeur la plus élevée, quel que soit le matériau considéré. Dans tous les cas, des molécules non identifiées à l'aide du scénario 6 apparaissent avec ce scénario.

Si, dans le cas particulier du « caoutchouc vulcanisé broyé », toutes les molécules identifiées à l'aide du scénario 6 sont présentes dans les résultats obtenus avec le scénario 5, pour les trois autres matériaux, ce n'est pas le cas.

On peut en particulier noter le cas de la molécule 37 dans le matériau « caoutchouc vulcanisé ». La valeur de seuil utilisée pour cette molécule dans le scénario 6 est entachée d'un faible niveau d'incertitude (valeur Labo). L'utilisation du seuil modélisé (scénario 5) ne permet pas de la retrouver, ce qui illustre, à ce stade, les limites d'une approche uniquement basée sur la modélisation.

La molécule 39 identifiée dans le « caoutchouc cru » et le « caoutchouc cru broyé » à l'aide du scénario 6 n'est pas présente dans les résultats obtenus avec le scénario 7. Dans le scénario 6, le seuil utilisé pour

cette molécule est issue de la base de données non nettoyée. La valeur utilisée présente donc potentiellement un niveau d'incertitude élevé. Dans le scénario 7, la valeur de seuil utilisée pour cette molécule est issue de la base de données nettoyée. L'effet du nettoyage est la disparition de cette molécule.

A l'inverse, ce même nettoyage fait apparaître la molécule 46 dans le matériau « caoutchouc vulcanisé broyé ». Il est intéressant de noter que cette molécule ressort également de l'utilisation des seuils modélisés (scénario 5). Malgré la limite évoquée ci-dessus, l'utilisation des valeurs modélisées apparaît donc, dans ce cas, pertinente.

Si l'on examine le cas des molécules 1 et 32, aucune valeur de seuil de perception n'est disponible pour elles dans la littérature. Le scénario 6 leur attribue donc un seuil égal à la moyenne géométrique des seuils de la famille chimique à laquelle elles appartiennent. Ces valeurs, caractérisées par un niveau d'incertitude très élevé, conduisent à ce que les molécules 1 et 32 ne ressortent jamais des résultats basés sur l'utilisation de ce scénario. A contrario, elles ressortent lors de l'utilisation des seuils modélisés (scénarios 5 et 7). La pertinence de ce résultat apparaît à l'examen de la nature chimique de ces deux molécules qui montre qu'elles appartiennent toutes les deux à des classes de composés connus pour être particulièrement odorants.

6.5 Conclusion

L'objectif visé dans cette partie de notre étude était de confronter la méthode développée à des données plus incertaines, car issues de l'analyse d'échantillons de matériaux pour lesquels l'intégralité des seuils de perception nécessaires à l'interprétation n'était pas disponible. Les résultats obtenus montrent clairement que le scénario 7, basé sur l'utilisation, pour chaque molécule identifiée, et par ordre de priorité, du seuil de perception mesuré au laboratoire, puis de la moyenne géométrique des seuils de perception disponibles dans la base de données nettoyée, et enfin la valeur issue de la modélisation locale si la molécule n'est pas renseignée dans la base, fournit des résultats non seulement très proches de ceux fournis par l'approche expert prise comme référence (scénario 6), mais les enrichit de l'identification de molécules supplémentaires, pertinentes au plan olfactif.

On peut donc conclure que l'outil mis au point dans le cadre de ce travail présente d'ores et déjà, à ce stade de développement, des performances intéressantes par la qualité des résultats qu'il fournit mais aussi par le gain de temps qu'il autorise par rapport à la démarche de l'expert.

Il convient toutefois de souligner que ces premiers résultats encourageants doivent être confirmés en analysant des matrices gazeuses présentant non seulement des dissimilitudes de composition plus importantes mais aussi des niveaux de concentration d'odeur plus élevés.

Enfin il conviendrait de vérifier par mesure olfactométrique la qualité des valeurs prédites par le modèle, ce qui permettrait de disposer de données complémentaires fiables pouvant être injectées dans le jeu de construction du modèle, et donc de le consolider.

CHAPITRE 7 : CONCLUSION GÉNÉRALE, DISCUSSION, ET PERSPECTIVES

7.1 Conclusion générale

L'objectif de ce travail de recherche était de mettre en place une méthodologie automatisée permettant d'identifier les molécules responsables de l'odeur d'une matrice gazeuse à partir des molécules présentes dans celle-ci. Ce travail visait en particulier à apporter une réponse à la problématique industrielle de l'identification des molécules à l'origine de l'odeur de matériaux.

Identifier les molécules à l'origine de l'odeur nécessite de croiser des données physicochimiques, des données olfactométriques et un troisième bloc de données : les valeurs de seuils de perception olfactive des molécules. Ces dernières données, informations clefs pour l'analyse d'un expert sont malheureusement caractérisées par des niveaux d'incertitude et d'incomplétude très élevés.

La première étape de notre travail a donc logiquement consisté à trouver des solutions pour pallier ces deux problèmes.

Un protocole permettant de fiabiliser les données de seuils de perception disponibles dans la littérature a été proposé. Il a comporté plusieurs étapes. La première a consisté à constituer une base de données initiale recensant, pour 650 molécules, des données de seuils de perception disponibles dans 165 articles. À partir de cette première base de données, une seconde a été construite en éliminant les valeurs issues des articles considérés comme *a priori* les moins fiables. Cela a conduit à l'élimination des articles proposant strictement moins de quatre valeurs de seuils de perception, c'est à dire plus de la moitié des articles, mais en conservant toutefois 97 % des molécules renseignées par au moins une valeur de seuil de perception.

Ensuite, pour chaque molécule, les valeurs aberrantes ont été éliminées à l'aide du test de Grubbs, ce qui a conduit à éliminer 10% du total des données de la base.

À l'étape suivante, un taux de comparaison et une fréquence d'exclusion de chaque article ont été définis. Seuls les articles ayant successivement un taux de comparaison strictement supérieur à 40% puis une fréquence d'exclusion inférieure ou égale à 12% ont été conservés. Cette dernière limite a été fixée à partir d'une classification ascendante hiérarchique. Cette étape a conduit à éliminer environ 50% de l'ensemble des données. Accepter un taux de perte d'information aussi élevé était justifié par la volonté de privilégier la fiabilité des données plutôt que leur nombre.

Au final, moins de la moitié des molécules renseignées par au moins une valeur de seuil de perception a été éliminée, le nombre de molécules renseignées dans la base ainsi nettoyée s'établissant à 338.

La dernière phase de ce processus a consisté en l'évaluation de la qualité du nettoyage. Pour cela, les valeurs de seuils de perception conservées ont été comparées à trente-trois seuils de perception mesurés par Olentica, ce bloc de données constituant notre référentiel. Dans vingt-huit cas (85%), une

amélioration de la valeur du seuil a été constatée après nettoyage. Dans trois cas (9%), il n'apporte ni amélioration, ni dégradation, et dans deux cas (6%), le scénario est strictement défavorable. Un test de significativité statistique a montré que le nettoyage permettait d'approcher la valeur expérimentale si l'on accepte un risque de 10%. Le nettoyage apparaît donc comme une alternative intéressante pour gérer le niveau élevé d'incertitude des données des seuils de perception disponibles dans la littérature.

L'étape suivante de notre travail a été ensuite de rechercher une solution pour gérer l'incomplétude de ces données. Pour cela, une méthodologie permettant de prédire les seuils de perception de molécules considérés comme non fiables ou non disponibles dans la littérature a été proposée. Cette modélisation se base sur un jeu de données consolidées de 350 molécules intégrant 45 seuils de perception mesurés par Olentica. Ces données d'entrée ont permis d'obtenir un modèle global satisfaisant, présentant une erreur de 1,06 sur le jeu de données de construction et 1,23 sur le jeu de données de validation (exprimé en logarithme décimal) et un coefficient de corrélation de 0,76 pour le jeu de données de construction et de 0,70 pour le jeu de données de validation.

Pour améliorer ce modèle global, un affinage par des modèles locaux par catégorie d'erreur a été appliqué. Pour cela, l'erreur de prédiction du seuil de perception de chaque molécule a été recensée pour permettre une catégorisation des molécules en fonction de cette erreur. Cet affinage a permis de diminuer l'erreur (exprimée en logarithme décimal) de 1,10 à 0,57 sur l'ensemble des données, ce qui apparaît comme très encourageant. En résumé, le modèle affiné, même s'il est clairement perfectible, permet d'obtenir des valeurs de seuils de perception correctes pour les molécules situées dans l'espace de construction. Lorsque la valeur de seuil de perception est inconnue, il montre de bonnes capacités de prédiction en fournissant des valeurs cohérentes avec les mesures réalisées par Olentica dans dix cas sur treize.

Des solutions aux problèmes liés à la fiabilité et à l'incomplétude des données de seuils de perception ayant été proposées, une méthodologie automatisée permettant d'identifier les molécules responsables de l'odeur, objectif principal de ce travail de recherche, a été mise en place. Cette méthodologie est principalement basée sur une automatisation du calcul des activités d'odeur de toutes les molécules présentes dans l'échantillon gazeux étudié. Cette automatisation permet non seulement un traitement exhaustif des données analytiques, mais aussi un gain de temps appréciable dans l'obtention des résultats par rapport à la démarche manuelle de l'expert.

Au-delà de cet aspect pratique, notre travail a visé à étudier dans quelle mesure le travail de l'expert pouvait être complété en apportant, par voie statistique, une information complémentaire sur des molécules pouvant présenter un impact significatif sur la variation de la concentration d'odeur d'un ensemble d'échantillons. Ce dernier point a nécessité un paramétrage des échantillons dont les résultats

pouvaient être à croiser. Il s'est avéré nécessaire de prendre en considération des études d'au moins quatre échantillons présentant un critère de similarité élevé.

L'ensemble de cette démarche a été développé à l'aide de données issues de l'analyse de mélanges gazeux synthétiques complexes (jusqu'à 48 molécules), ce qui a permis de maîtriser les incertitudes liées à l'identification, à la quantification, et aux seuils de perception, l'essentiel de ces derniers ayant été mesurés par Olentica.

La suite de notre travail s'est attachée à évaluer l'impact de la qualité des données d'entrée (précision des concentrations physico-chimiques et sensorielles, précision des seuils de perception) sur la qualité des résultats de la démarche. À l'issue de cette partie de l'étude, plusieurs conclusions peuvent être dégagées pour chacune des deux approches, quant à l'impact du niveau de précision des données utilisées sur la détermination des molécules responsables de l'odeur.

Concernant l'approche « expert », il ressort que l'amélioration de la quantification fait que les activités d'odeur calculées sont plus représentatives du poids réel des molécules dans la concentration d'odeur du mélange. Pour les deux mélanges étudiés, seule la quantification absolue permet d'obtenir une identification et une hiérarchisation conformes à la référence.

Parmi l'ensemble des scénarios envisagés pour obtenir les seuils de perception, la modélisation locale ressort comme étant celui qui permet d'approcher au mieux les résultats de l'expert.

Enfin, la précision sur la mesure olfactométrique n'apparaît pas avoir d'influence significative sur l'identification et la hiérarchisation des molécules responsables de l'odeur. La concentration d'odeur mesurée lui sert uniquement de référence pour s'assurer de la pertinence de son identification.

Concernant l'approche statistique, la PLS fait ressortir les molécules expliquant au mieux la variation de la concentration d'odeur de l'ensemble des données utilisées, et non spécifiquement celles responsables de l'odeur du mélange étudié. Si une amélioration de la précision des données d'entrée (quantifications physicochimiques et olfactométriques, seuils de perception) doit logiquement conduire à une meilleure qualité de résultat, cette amélioration s'avère difficilement discernable dans le cadre de notre étude.

Enfin, il convient de souligner que ces résultats ont été obtenus pour des mélanges étalons constitués d'une proportion très importante (>75%) de molécules dont les seuils de perception avaient été mesurés dans le cadre de cette étude. Une telle proportion sera clairement inatteignable lors d'une analyse courante, le nombre de seuils de perception disponibles dans la littérature étant très faible au regard du nombre de molécules susceptibles d'être identifiées.

L'objectif visé dans la dernière partie de notre étude était donc de confronter la méthode développée à des données plus incertaines, issues de l'analyse d'échantillons de matériaux pour lesquels la majorité

des seuils de perception nécessaires à l'interprétation n'était pas disponible. Nos résultats ayant montré que la modélisation locale permettait d'approcher au mieux les résultats de l'expert, il nous est donc apparu pertinent d'envisager à ce stade un nouveau scénario les combinant. Le nouveau scénario ainsi élaboré a consisté, pour chaque molécule identifiée, à utiliser prioritairement le seuil de perception mesuré au laboratoire, puis la moyenne géométrique des seuils de perception disponibles dans la base de données nettoyée, et enfin la valeur issue de la modélisation locale si la molécule n'était pas renseignée dans la base

Les résultats obtenus font que ce scénario fournit des résultats non seulement très proches de ceux fournis par l'approche « expert » prise comme référence, mais les enrichit de l'identification de molécules supplémentaires, pertinentes au plan olfactif.

On peut donc conclure que l'outil mis au point dans le cadre de ce travail présente, en l'état, des performances intéressantes par la qualité des résultats qu'il fournit mais aussi par le gain de temps qu'il autorise par rapport à la démarche manuelle de l'expert.

Il convient toutefois de souligner que ces premiers résultats encourageants doivent être confirmés en analysant des matrices gazeuses présentant non seulement des dissimilitudes de composition plus importantes mais aussi des niveaux de concentration d'odeur plus élevés.

Enfin il conviendrait de vérifier par mesure olfactométrique la qualité des valeurs de seuils de perception prédites par le modèle, ce qui permettrait de disposer de données complémentaires fiables pouvant être injectées dans le jeu de construction du modèle, et donc de le consolider.

7.2 Discussions

Après avoir présenté les résultats de cette recherche, la suite de cette conclusion générale propose de discuter des apports et des limites de ce travail, en les mettant en perspective avec les développements futurs qu'il serait pertinent d'envisager.

7.2.1 Les concepts

La première partie de ce travail a permis de présenter un aperçu des connaissances actuelles dans le domaine de l'olfaction. Nous nous sommes intéressés aux divers facteurs influençant la perception olfactive à travers un état de l'art qui a montré l'importance de la physiologie, de la cognition et de l'état émotionnel d'un sujet sur la perception des odeurs. Cet état de l'art a permis de souligner le caractère très subjectif cette perception chez l'homme.

Pour contourner cette difficulté, nous avons choisi de limiter notre étude à une des caractéristiques de l'odeur : sa persistance. L'évaluation de ce paramètre présente en effet l'avantage d'être la seule à être réalisée à des niveaux de concentration dits liminaires. A ces niveaux, la qualité de l'odeur n'est pas perceptible par les jurés, leur réponse est donc binaire, non entachée par la perception du caractère

agréable ou désagréable de l'odeur. Elle est donc susceptible d'être traitée statistiquement et présente des niveaux d'incertitude beaucoup plus faibles que ceux liés à l'évaluation des autres caractéristiques de l'odeur.

Sur un autre plan, le concept « d'expert » a été présenté. L'expert a été défini de manière générale comme une personne maîtrisant la connaissance dans le ou les domaines utiles à la réalisation de l'expertise et possédant une expérience reconnue qui lui permet d'évaluer les éléments faisant partie de son champ d'expertise. Il s'agit donc là d'une notion particulièrement floue.

La variabilité des résultats des experts, si elle est évidemment largement liée à l'incertitude des données qu'ils manipulent, est le plus souvent principalement liée à la disparité des compétences des experts eux-mêmes. Dans le contexte qui nous concerne, l'expert doit *a minima* maîtriser la métrologie des gaz et celle des odeurs, et de préférence posséder une solide connaissance de la nature des matrices gazeuses issues des différents secteurs industriels, des molécules, et de leurs propriétés odorantes.

A compétences égales, la qualité du résultat de l'analyse de l'expert dépend en particulier de la finesse de son analyse, c'est-à-dire de sa capacité, ou de sa volonté, à faire porter son analyse jusqu'aux molécules minoritaires, présentes à l'état de traces dans la matrice analysée.

A ce niveau, l'outil développé dans ce travail apporte une réponse au problème en uniformisant à la fois la qualité et l'exhaustivité du traitement de l'information disponible.

Enfin, un concept largement manipulé dans le cadre de cette étude est celui d'incertitude. Sur un plan tout à fait général, cette incertitude gêne l'évaluation et la résolution des problèmes complexes. Ses principales dimensions sont techniques (l'inexactitude), méthodologiques (le manque de fiabilité), épistémologique (l'ignorance). L'inexactitude fait référence à la précision numérique des données qui, comme nous l'avons vu, peut être liée à la capacité de résolution limitée des appareils de mesure.

Le manque de fiabilité est lui associé aux limites des méthodes que nous employons. Si par exemple un seuil de perception est obtenu, comme nous l'avons réalisé, par une méthode d'estimation, il est moins fiable que s'il était fondé sur un grand nombre de mesures directes. Enfin, l'incertitude épistémologique trouve sa source dans notre capacité limitée à savoir et à comprendre. Elle intervient lorsqu'un expert se rend compte des champs dans lesquels sa connaissance des odeurs est limitée, mais n'a aucune idée de l'importance relative de ces champs. Il est donc incapable de les inclure correctement dans son analyse.

Il faut donc bien considérer que les techniques statistiques quantitatives que nous avons largement utilisées, si elles sont essentielles dans toute analyse de l'incertitude, ne fournissent qu'une perception partielle d'un ensemble complexe d'incertitudes. Dans de nombreux cas en effet, les dimensions liées au manque de fiabilité ou à l'ignorance peuvent dominer l'incertitude globale. Cela souligne donc la portée limitée de ces méthodes.

7.2.2 Les apports de la démarche

Pour mettre en avant les apports de nos travaux, deux domaines ont été définis : Les contributions méthodologiques qui peuvent constituer des bases pour des recherches futures, et les contributions opérationnelles qui attestent de l'applicabilité de la méthode développée.

7.2.2.1 Les apports méthodologiques

La problématique de l'identification des molécules à l'origine d'une l'odeur est un axe de recherche qui est appelé à se développer car il est sous-tendu par une exigence croissante des populations et des consommateurs pour une maîtrise des odeurs auxquelles ils sont confrontés, qu'il s'agisse d'odeurs environnementales ou liées aux produits qu'ils utilisent. Cette exigence devient de plus en plus structurante pour l'activité industrielle au sens large, et pour celle des matériaux en particulier. Dans ce cadre général, les apports méthodologiques de ce travail se situent sur plusieurs plans.

Le premier apport concerne l'utilisation des valeurs de seuils de perception olfactive des molécules renseignées dans la littérature, valeurs caractérisées par un niveau élevé d'incertitude et d'incomplétude. Une méthodologie d'amélioration de l'utilisation de ces données a été proposée. Elle se décline en trois étapes :

- Une fiabilisation des données disponibles de manière à éliminer les mesures de seuils de perception pouvant être considérées comme statistiquement les moins fiables (nettoyage de la base de données initiale) ;
- Une modélisation globale des seuils de perception retenus après cette étape de nettoyage. Cette modélisation est réalisée sur la base d'une approche QSAR/QSPR. Elle permet de pallier l'incomplétude des données de seuils de la littérature ;
- Enfin, une modélisation locale complémentaire basée sur une analyse des erreurs du modèle global.

Un second apport méthodologique a consisté à maîtriser la variabilité et la « granularité » de l'analyse de l'expert en proposant une automatisation de la démarche d'identification des molécules responsables de l'odeur basée sur l'utilisation des seuils nettoyés et de ceux prédits par le modèle élaboré.

7.2.2.2 Les apports opérationnels

Les apports opérationnels sont étroitement liés aux contributions méthodologiques qui viennent d'être citées.

En effet, les résultats obtenus sur le plan méthodologique ont permis de mettre au point un outil permettant d'identifier automatiquement les molécules responsables de l'odeur d'un mélange gazeux. Cet outil présente plusieurs avantages. En premier lieu, il permet de se prémunir de la variabilité

d'analyse liée à l'expert, mais il se traduit par un gain de temps par rapport à ce dernier, et il apporte une certaine autonomie à l'industriel face à l'analyse d'une problématique d'odeur.

La mise en place de cet outil a en outre permis de définir un cadre précis d'utilisation des données sources (seuils de perception, données physico-chimiques et sensorielles).

La connaissance des molécules responsables des nuisances odorantes ainsi rendue possible constitue une aide à la décision pour des industriels. Plus précisément, cette approche peut s'avérer utile en conditions opérationnelles dans deux contextes particuliers :

- dans le cadre d'une problématique d'odeurs de matériaux,
- dans celui de la gestion des émissions gazeuses d'un site émetteur.

Dans ces deux contextes, l'identification des molécules responsables de l'odeur revêt en effet une importance stratégique. Dans le cas des matériaux, l'odeur est de nos jours de plus en plus susceptible de déclencher ou d'inhiber la décision d'achat ou d'utilisation d'un produit. Dans le cas de la gestion des émissions gazeuses de sites industriels, l'odeur est là aussi le premier signal que reçoit le riverain sur la qualité de l'air qu'il respire, et donc le premier motif de plainte, pour peu que l'odeur perçue ait un caractère désagréable, ou « inconnu », qui le lui fera immédiatement associer à un potentiel risque pour la santé. Dans ces deux cas, l'identification des molécules responsables de l'odeur est donc fondamentale pour prendre les bonnes décisions, qu'il s'agisse de définir la bonne formulation pour les matériaux, ou le procédé de traitement pour les émissions industrielles.

Au delà de ces aspects, un des bénéfices de la démarche mise au point dans le cadre de ce travail est que la claire identification des molécules responsables de l'odeur permet de découpler les notions d'odeur et de toxicité, et donc de rationaliser le discours en ne mélangeant pas les concepts.

7.3 Limites et perspectives de la recherche

Notre travail de recherche avait une visée opérationnelle clairement revendiquée. Il convient donc de définir les limites des résultats obtenus, limites qui constituent par ailleurs autant de perspectives d'amélioration pour le futur.

Quatre axes limitant principaux peuvent être identifiés:

- L'identification et la quantification des molécules présentes dans un mélange gazeux
- La gestion des données de seuils de perception existantes
- La modélisation des seuils de perception
- L'outil lui-même

7.3.1 L'identification et la quantification des molécules

Nos résultats, en particulier ceux issus des expérimentations conduites sur de mélanges synthétiques, ont permis de montrer que plus la connaissance des caractéristiques de l'ensemble des molécules présentes dans la matrice gazeuse étudiée était précise, tant au plan qualitatif (nature chimique des molécules) que quantitatif (concentration des molécules), meilleur était, à qualité de seuils de perception égale, le résultat de l'identification des molécules responsables de l'odeur.

Il apparaît donc que l'utilisation conjointe de techniques séparatives présentant des niveaux de sensibilité et de sélectivité élevés, comme la chromatographie en phase gazeuse bidimensionnelle couplée à la spectrométrie de masse, et d'une technique permettant une quantification rapide comme la spectrométrie de masse à résonance protonique, permettrait certainement d'envisager un gain substantiel sur les deux plans cités ci-dessus, et donc sur le résultat global.

7.3.2 La gestion des données de seuils de perception existantes

Comme cela a été largement souligné dans ce travail, le seuil de perception est l'élément clé de l'identification des molécules responsables de l'odeur d'un mélange gazeux. La gestion de ces données est donc fondamentale. Si le protocole de nettoyage statistique que nous avons mis en place a permis d'obtenir des résultats satisfaisants, il faut souligner qu'il se traduit par un niveau de perte d'information important. Seules 338 molécules sur les 650 initiales restent en effet renseignées par au moins une valeur de seuil de perception après nettoyage de la base de données. Ce nombre est faible surtout lorsqu'on le compare aux milliers de molécules susceptibles d'être à l'origine d'odeurs, ce qui pose par ailleurs avec d'autant plus d'acuité le problème de la gestion de l'incomplétude de ces données.

Deux axes d'amélioration sont envisageables à ce niveau. Tout d'abord, augmenter la taille de la base de données en mettant en place une veille scientifique automatisée qui permette de nourrir la base de données initiale avec les valeurs régulièrement publiées. L'autre axe consiste à améliorer le processus de nettoyage des données. Notre démarche a en effet été construite autour de l'utilisation d'un taux de comparaison des valeurs de seuils obtenues pour une même molécule par différents auteurs. Cette approche, si elle a permis de fiabiliser significativement les données, a eu toutefois pour conséquence de nous conduire à éliminer des valeurs de seuils au titre qu'elles ne pouvaient être comparées au reste du jeu de données, même si elles provenaient à l'évidence d'un auteur dont les résultats pouvaient être jugés fiables.

Une perspective d'amélioration serait d'attribuer dès le début de la démarche un poids à chacun des articles pour faire une moyenne en fonction de leur « fréquence d'exclusion » et de leur « taux de comparaison » afin de conserver toute l'information initiale. Cela permettrait de pondérer l'information à l'aide de fonctions de croyance pour mieux représenter l'incertitude de base tout en la conservant et donc mieux gérer le sur-apprentissage.

7.3.3 La modélisation des seuils de perception

L'amélioration de la modélisation des seuils de perception apparaît clairement comme une des pistes majeures pour la suite de ce travail. C'est en effet la voie qui semble la plus réaliste pour faire face à l'incomplétude des données actuelles. L'autre voie qui permettrait de pallier ce problème serait la mesure expérimentale des seuils manquants, mais cette approche, si elle reste fondamentalement nécessaire, ne peut que constituer une solution marginale. Elle se heurte en effet à plusieurs écueils : la mesure du seuil de perception d'une molécule doit être répétée un nombre significatif de fois pour pouvoir lui attribuer un intervalle de confiance, c'est donc un processus lourd, consommateur de temps, et par là même, coûteux. De plus, la mesure est conditionnée par la possibilité de se fournir la molécule dont on souhaite évaluer le seuil à un niveau de pureté suffisant, ce qui s'avère fréquemment impossible. Enfin, comme souligné précédemment, la diversité des molécules susceptibles d'être à l'origine de problèmes d'odeur est telle, qu'indépendamment de leur disponibilité, il est illusoire d'envisager de gérer l'incomplétude des données sur la simple base de mesures.

Le modèle que nous avons élaboré, affiné grâce à une modélisation locale, même s'il est clairement perfectible, s'avère satisfaisant dans la mesure où il permet d'obtenir une valeur de seuil de perception correcte pour les molécules situées dans l'espace de construction, même si l'on note un sur-apprentissage marqué.

Afin de limiter le sur-apprentissage de la phase d'affinage, une des pistes d'amélioration serait de proposer, au même titre que pour le modèle global, une constitution du jeu de construction couvrant à la fois l'espace des structures des molécules et la représentativité des seuils de perception, et non pas uniquement cette dernière.

Une autre piste d'amélioration, présentée au paragraphe précédent, serait d'attribuer dès le début de la démarche un poids à chacun des articles pour en faire une moyenne en fonction de leur fréquence d'exclusion et de leur taux de comparaison. Cela permettrait en effet de conserver toute l'information initiale et de la pondérer à l'aide de fonctions de croyance pour mieux représenter l'incertitude de base tout en la conservant et donc de mieux gérer le sur-apprentissage (P. XU, DAVOINE, et DENOEU 2014).

Enfin, une amélioration des modèles locaux passerait par une meilleure sélection du nombre de composantes lors de l'analyse en composantes indépendantes.

Concernant la sélection des indices de structure qui ont servi à la construction du modèle, ils ont été choisis en fonction de leur intérêt dans l'explication de l'odeur et de la possibilité de les calculer à partir du SMILES des molécules. Cependant, tous les indices potentiellement utilisables n'ont pas le même degré de complexité de calcul et ne nécessitent donc pas le même temps de programmation. Un compromis a donc été fait dans le cadre de notre étude, ce qui en constitue donc une des limites. Une

piste d'amélioration pourrait donc être l'intégration d'autres indices, parmi lesquels on peut citer, sans prétention d'être exhaustif :

- L'indice de Lu (2006) pour traduire l'électronégativité de la molécule,
- L'indice Epsilon d'Ernesto et Estrada de 1994 pour traduire l'isométrie,
- Un indice évaluant la symétrie de la molécule (IVANCIUC 1998),
- L'indice DAI (2006) pour le caractère spatial,
- L'indice de connectivité mF (MU, FENG, et XU 2006)

Enfin, on peut rappeler ici qu'une perspective d'amélioration de la modélisation pourrait consister à vérifier par mesure olfactométrique la qualité des valeurs prédites par le modèle, surtout dans le cas de molécules pour lesquelles l'incertitude est importante (molécules polyfonctionnelles) ce qui permettrait de disposer de données complémentaires fiables pouvant être injectées dans le jeu de construction du modèle, et donc de le consolider.

7.3.4 L'outil d'identification des molécules responsables de l'odeur

Les résultats que nous avons obtenus montrent que l'outil d'identification des molécules responsables de l'odeur fonctionne de façon très satisfaisante dans un cas théorique, lorsque les incertitudes sur les différents blocs de données sont bien maîtrisées, mais qu'il nécessite davantage de précautions, et donc d'adaptations, lorsque la qualité des données de base est dégradée.

La principale limite de la PLS qui a été utilisée dans le cadre de notre travail est qu'en attribuant même poids (VIP) à toutes les molécules quel que soit l'échantillon lorsque les échantillons sont analysés ensemble, elle fait ressortir les molécules expliquant au mieux la variation de la concentration d'odeur de l'ensemble des données utilisées, et non spécifiquement celles responsables de l'odeur du mélange étudié. Une piste d'amélioration pourrait à ce niveau consister à déterminer une limite de sélection VIP en fonction de l'incertitude des données utilisées.

Sur un plan général, il faut rappeler ici qu'une limite actuelle de notre approche est qu'elle s'est volontairement limitée à rechercher les corrélations entre les concentrations des molécules présentes, leurs seuils de perception et la concentration d'odeur du mélange gazeux. Les autres caractéristiques de l'odeur : son intensité, sa qualité, son acceptabilité, n'ont pas été prises en compte. Cela constitue une des perspectives majeures d'évolution de cet outil, mais aussi un défi d'envergure, eu égard là encore à l'incertitude et l'incomplétude de ces blocs de données, ainsi qu'à la complexité de l'évaluation de ces caractéristiques.

TABLE DES FIGURES

Figure 1 : Anatomie du système olfactif (GANTY 2016).....	20
Figure 2 : Structure de la muqueuse olfactive	21
Figure 3 : Mécanisme de l'olfaction (GANTY 2016).....	21
Figure 4 : Courbe de Stevens	24
Figure 5 : Odeur persistante/ odeur fugace.....	25
Figure 6 : Évolution de l'intensité d'une odeur en fonction de la distance à la source émettant l'odeur	25
Figure 7 : Schéma de principe d'un olfactomètre à dilution dynamique (Doc Olentica®).....	28
Figure 8 : Olfactomètre à dilution dynamique (ODILE®).....	28
Figure 9 : Ampoule de prélèvement	31
Figure 10 : Canister	31
Figure 11 : Sacs d'échantillonnage.....	32
Figure 12 : Détecteur à photoionisation (PPB RAE System).....	36
Figure 13 : Structure du méthoxyacétaldéhyde	39
Figure 14 : Format de la base de données des seuils de perception recensés	46
Figure 15 : Quantité d'information disponible dans BDD1 (par ordre décroissant du nombre de valeurs par molécule sur l'axe des abscisses et par ordre décroissant du nombre de valeurs par article sur l'axe des ordonnées).....	47
Figure 16 : Fréquence du nombre de valeurs seuils de perception mesurées par molécule	50
Figure 17 : Distribution des valeurs de seuils de perception des 11 molécules les plus renseignées dans la littérature après l'étape 2	51
Figure 18 : Diagramme Q-Q du n-butanol (CAS=71-36-3).....	52
Figure 19 : Coefficient de corrélation (CC) des diagrammes Quantile-Quantile (Q-Q) des molécules dont le seuil de perception est renseigné par strictement plus de 10 articles après l'étape 2 du nettoyage	53
Figure 20 : Dendrogramme de la CAH appliquée sur les FE des articles ayant un TC strictement supérieur à 40%	58
Figure 21 : Évolution de la perte d'inertie (en pourcentage) en fonction du nombre de classes	59
Figure 22 : Évolution du SPRSQ en fonction du nombre de classes	59
Figure 23 : HeatMap des valeurs de seuils de perception du référentiel standardisées par molécule ...	64
Figure 24 : HeatMap des valeurs de seuils de perception du référentiel standardisées par molécules après application du test de Dixon.....	65
Figure 25 : Comparaison de Z'_1 , Z'_2 et Re' (en noir).....	67
Figure 26 : Illustration de l'InChI et du SMILES pour l'indole	78
Figure 27 : HeatMap de la répartition des familles chimiques en fonction des combinaisons de groupements fonctionnels.....	81
Figure 28: Classification CAH des molécules consolidée par l'indice de Hopkins.....	85
Figure 29: Projection des molécules dans l'espace défini par les axes principaux de la méthode de « Projection Pursuit »	88
Figure 30: Projection des indices dans l'espace défini par les axes principaux de la méthode de Projection Pursuit	89
Figure 31 : Projection des molécules dans l'espace défini par les axes principaux de la Projection Pursuit.....	91
Figure 32 : (a) Histogramme de la composition des classes en familles chimiques (b) et répartition des familles chimiques dans ces classes	93

Figure 33 : Projection des molécules par classe sur l'espace défini par les axes principaux de la « Projection Pursuit »	95
Figure 34 : Division du jeu de données en données de construction (« train ») et en données de validation (« test ») pour la classe 3 (96 molécules)	96
Figure 35 : Division du jeu de données en données de construction (« train ») et en données de validation (« test ») pour la classe 2 (8 molécules)	97
Figure 36 : Division des 350 molécules en données de construction (« train ») et en données de validation (« test »).....	97
Figure 37 : Valeurs des seuils de perception à modéliser dans l'ordre croissant	98
Figure 38 : Valeurs de seuils de perception (en log10) à modéliser classées dans l'ordre croissant	99
Figure 39 : Données à mettre en relation pour le modèle global.....	99
Figure 40 : Erreurs calculées sur le jeu de construction (« train ») et sur le jeu de validation (« test ») en fonction de la valeur d'épsilon	102
Figure 41 : Graphe des valeurs prédites en fonction des valeurs observées (350 molécules).....	103
Figure 42 : Projection des molécules sur les axes de « Projection Pursuit » selon l'erreur de prédiction de leur seuil de perception par le modèle global (arrondie à l'unité près)	107
Figure 43 : Répartition des molécules sur l'axe 1 de « Projection Pursuit » selon l'erreur de prédiction par le modèle global	108
Figure 44 : Description des erreurs de prédiction des seuils de perception par le modèle global en fonction des groupements fonctionnels des molécules.....	109
Figure 45 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 1, avant et après affinage par modélisation locale	111
Figure 46 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 2, avant et après affinage par modélisation locale	112
Figure 47 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 3, avant et après affinage par modélisation locale	113
Figure 48 : Comparaison de la prédiction des seuils de perception dont l'erreur de prédiction par le modèle global se situe dans la catégorie 4, avant et après affinage par modélisation locale	114
Figure 49 : Comparaison de la prédiction des seuils de perception des 350 molécules avant et après affinage par modélisation locale.....	116
Figure 50 : Projection des molécules dont le seuil de perception n'est pas renseigné dans la littérature dans l'espace défini par les axes principaux de « Projection Pursuit » obtenus à partir des molécules ayant au moins un seuil de perception renseigné	123
Figure 51 : Positionnement des molécules dans l'espace de justesse et précision.....	130
Figure 52 : Plan d'étude de la répétabilité de la fabrication d'un mélange gazeux synthétique	132
Figure 53 : Format des données utilisées dans l'approche statistique.....	138
Figure 54 : Proportion de l'activité d'odeur de chaque molécule du mélange 5 par rapport à la somme des activités d'odeur de ce mélange	142
Figure 55 : Evolution de la RMSEP en fonction du nombre de composantes PLS utilisé.....	143
Figure 56 : VIP des molécules de l'étude menée sur le mélange 5 en référence et un critère de similarité de 4000	144
Figure 57 : Comparaison quantification en équivalent toluène / concentrations expérimentales pour le mélange F	151
Figure 58 : Comparaison quantification en équivalent toluène / concentrations expérimentales pour le mélange G	152
Figure 59 : Erreur relative (%) moyenne de la quantification en équivalent toluène par rapport à la concentration expérimentale sur les mélanges F et G	152

Figure 60 : Comparaison quantification en équivalent famille / concentrations expérimentales pour le mélange F	153
Figure 61 : Comparaison quantification en équivalent famille / concentrations expérimentales pour le mélange G	154
Figure 62: Erreur relative (%) moyenne de la quantification en équivalent famille par rapport à la concentration expérimentale sur les mélanges F et G	154
Figure 63 : Comparaison quantification absolue / concentrations expérimentales pour le mélange F	155
Figure 64 : Comparaison quantification absolue / concentrations expérimentales pour le mélange G	155
Figure 65 : Erreur relative (%) de la quantification absolue par rapport à la concentration expérimentale sur les mélanges F et G	156
Figure 66 : Concentrations d'odeur expérimentales (COexp) et théoriques (COth)	163
Figure 67 : Sacs en Nalophan®	176
Figure 68 : Olfactomètre à Dilution Dynamique Odile®	177
Figure 69 : Étapes du décodage de la grandeur SMILES de l'indole.....	231
Figure 70 : Matrices utilisées pour le calcul des indices topologiques - Cas des molécules avec liaisons simples et sans hétéroatomes	232
Figure 71 : Matrices étendues utilisées pour construction des indices topologiques - Cas des molécules avec liaisons tous types de liaisons et avec hétéroatomes	232
Figure 72 : Définition de la matrice CEP (« weighted electronic connectivity matrix »)	233

TABLE DES TABLEAUX

Tableau 1 : Classification du ton hédonique (VDI-3882-2 1994)	26
Tableau 2 : Exemples de seuils de perception d'après « Pollutions Olfactives » de l'ADEME (ADEME 2005).....	29
Tableau 3 : Variabilité des valeurs de seuils de perception de l'acétate de butyle (VAN GEMERT 2011).....	43
Tableau 4 : Distribution des articles en fonction du nombre de valeurs de seuils de perception mesurés	48
Tableau 5 : Bilan de la perte d'information de l'étape 2 du nettoyage	49
Tableau 6 : Molécules renseignées par un minimum de 10 valeurs de seuil de perception	51
Tableau 7 : Exemple 1 : molécules renseignées une seule fois (valeurs de seuils de perception en mg/m3)	54
Tableau 8 : Exemple 2 : molécules renseignées une seule fois (valeurs de seuils de perception en mg/m3)	54
Tableau 9 : Bilan de la perte d'information à l'étape 3 du nettoyage	55
Tableau 10 : Élimination de valeurs de seuils de perception : détail de la perte d'information.....	55
Tableau 11 : Articles éliminés ($TC \leq 40\%$).....	57
Tableau 12 : Bilan de la perte d'information après élimination des articles dont le $TC \leq 40\%$	57
Tableau 13 : Bilan de la perte d'information après élimination des articles (CAH sur les FE).....	60
Tableau 14 : Bilan de la perte d'information de l'étape 4 du nettoyage	60
Tableau 15 : Mesures des seuils de perception pour 45 molécules du référentiel.....	61
Tableau 16 : Définition des valeurs calculées par molécule pour comparaison des jeux de données... 63	63
Tableau 17 : Valeurs de seuils de perception (SP) éliminées par le test de Dixon.....	65
Tableau 18 : Définition des variables à comparer après transformation	66
Tableau 19 : Scénarios consécutifs au nettoyage	67
Tableau 20 : Molécules correspondant aux trois scénarios négatifs (5,7, et 8)	68
Tableau 21 : Variables définies par molécule	69
Tableau 22 : Molécules dont toutes les valeurs de seuils de perception ont été éliminées par le nettoyage	69
Tableau 23 : Bilan de la perte d'information totale liée au nettoyage	70
Tableau 24 : Résumé d'études statistiques portant sur la prédiction du seuil de perception de molécules (RN = Réseau de Neurones ; RL= régression Linéaire ; RLM = Régression Linéaire Multiple ; PLS = « Partial Least square Regression » ; SVM=Support Vecteur Machine ;	76
Tableau 25 : Liste des variables d'entrée et leur notation.....	79
Tableau 26 : Tri des indices par la technique du Lasso.....	87
Tableau 27 : Indice de Hopkins de la fusion des classes deux par deux	92
Tableau 28 : Résultat des erreurs par validation croisée	104
Tableau 29 : Intervalles des erreurs (log10) permettant de décomposer le jeu de données de molécules en sous-domaines pour l'affinage local du modèle global.	105
Tableau 30 : Répartition des molécules par catégorie pour la modélisation locale de l'erreur.....	110
Tableau 31 : Comparaison de la modélisation globale et de la modélisation globale additionnée de l'affinage sur les molécules dont le seuil de perception a été mesuré par Olentica	117
Tableau 32 : Comparaison du résultat de la modélisation globale (avec et sans affinage local) avec les valeurs expérimentales d'Olentica (valeurs d'entrée du modèle = valeurs de la littérature nettoyée) 119	119
Tableau 33 : Molécules situées hors de la couverture de recouvrement des molécules renseignées ..	124

Tableau 34 : Molécules et seuils de perception utilisés pour la constitution des 5 mélanges gazeux synthétiques	128
Tableau 35 : Composition des cinq mélanges liquides synthétiques (masses pesées (mg))	131
Tableau 36 : Proportion de seuils de perception mesurés expérimentalement par mélange	131
Tableau 37 : Évaluation des incertitudes lors de la constitution des mélanges gazeux	133
Tableau 38 : Composition massique des 5 mélanges gazeux synthétiques (mg/m3)	133
Tableau 39 : Activités d'odeur des molécules pour les cinq mélanges gazeux synthétiques (en grisé, les activités d'odeur inférieure à 1)	134
Tableau 40 : Comparaison des concentrations d'odeur mesurées avec la somme des activités d'odeur (AO_{som}) et l'activité d'odeur maximale (AO_{max})	135
Tableau 41 : Composition et activités d'odeur des mélanges virtuels	136
Tableau 42 : Caractéristiques des mélanges virtuels	137
Tableau 43 : Molécules responsables de l'odeur de chaque mélange selon l'expert	137
Tableau 44 : Critères de similarité du mélange 5 avec les autres mélanges	143
Tableau 45 : Concentration d'odeur expérimentale des échantillons de l'analyse	143
Tableau 46 : Pourcentage de variance de Y expliqué par les composantes orthogonales « t » de PLS	144
Tableau 47 : Comparaison de l'approche de l'expert et de l'approche statistique pour différentes valeurs de critère de similarité	145
Tableau 48 : Concentrations d'odeur des différentes études pour un critère de similarité de 2500	146
Tableau 49 : Calcul du ΔZ des différentes études tel que défini dans la norme EN13725	146
Tableau 50 : Critères de similarité du mélange 3 avec les autres mélanges	147
Tableau 51 : Pourcentages d'erreur relative par rapport à la concentration expérimentale	156
Tableau 52 : Molécules identifiées comme responsables de la concentration d'odeur suivant le mode de quantification et l'approche (mélanges F et G)	158
Tableau 53 : Logigramme d'utilisation des seuils de perception	161
Tableau 54 : Valeurs des seuils de perception obtenues grâce aux différents scénarios	162
Tableau 55 : Concentrations d'odeur expérimentales (CO_{exp}) et théoriques (CO_{th})	163
Tableau 56 : Erreurs relatives pour les différents mélanges et scénarios	164
Tableau 57 : Parallèle entre la concentration des molécules dans les mélanges et valeur du seuil de perception selon le scénario	164
Tableau 58 : Molécules identifiées comme responsables de la concentration d'odeur suivant le scénario de seuils de perception utilisé et l'approche (mélanges 2 et 5)	166
Tableau 59 : Comparaison des valeurs de concentration d'odeur (UOE/m3) obtenues avec les jurys certifiés (6 et 12 jurés)	169
Tableau 60 : Comparaison des valeurs de concentration d'odeur (UOE/m3) obtenues avec les jurys non certifiés (6 et 12 jurés)	169
Tableau 61 : Comparaison des valeurs de concentration d'odeur (UOE/m3) obtenues avec les jurys non certifiés (24 jurés)	169
Tableau 62 : Comparaison des taux de variation des concentrations d'odeur (UOE/m3) obtenues avec les différents types de jurys	170
Tableau 63 : Approche statistique : Molécules identifiées comme responsables de la concentration d'odeur (mélanges 2 et 5)	171
Tableau 64 : Matériaux analysés	176
Tableau 65 : Nombre de molécules présentes dans les émissions des matériaux	178
Tableau 66 : Concentrations des molécules présentes dans les différents échantillons	179
Tableau 67 : Concentration d'odeur des échantillons	180

Tableau 68 : Valeurs des seuils de perception obtenues suivant les scénarios.....	180
Tableau 69 : Identification des molécules responsables de l'odeur : résultats de l'approche expert automatisée.....	182
Tableau 70 : Concentrations des molécules du mélange 1 synthétique à l'état liquide et gazeux	273

BIBLIOGRAPHIE

- ABRAHAM, Michael H., Joëlle M.R. GOLA, J. Enrique COMETTO-MUNIZ, et William S. CAIN. 2002. « A Model for Odour Thresholds ». *PubMed* 27 (2): 95 - 104.
- ADALI, Tulay, Christian JUTTEN, Joao Marcos TRAVASSOS ROMANOMARCOS, et Allan Kardec BARROS. 2009. « Factor Analysis as Data Matrix Decomposition ». In *Independent Component Analysis and Signal Separation: 8th International Conference*, Springer, 785. ICA 2009, Paraty, Brazil, March 15-18, 2009, Proceedings. Paraty, Brésil: Springer.
<https://books.google.fr/books?id=ZtdsCQAAQBAJ&pg=PA168&lpg=PA168&dq=mixing+matrix+loading&source=bl&ots=l6kEi613i9&sig=nVtGjjJNdHZhADU2rMlmlIAHL0zs&hl=fr&sa=X&ved=0ahUKEwidgKnIs43cAhUiQpoKHcYcC1UQ6AEIZTAQ#v=onepage&q=mixing%20matrix%20loading&f=false>.
- ADEME. 2005. *Pollutions olfactives - L'usine Nouvelle*. Dunod. Paris.
- ADOLFSSON, Andreas, Margareta ACKERMAN, et Naomi C. BROWNSTEIN. 2017. « To Cluster, or Not to Cluster: How to Answer the Question ». In *Proceedings of Knowledge Discovery from Data*, 9. Halifax, Nova Scotia, Canada: Proceedings of Knowledge Discovery from Data.
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>.
- AFNOR. 1996. *Norme NF X43-103 : Qualité de l'air. Mesurage de l'odeur d'une atmosphère gazeuse. méthode supraliminaire*. <http://sagaweb.afnor.org/fr-FR/sw/Consultation/Pdf/1259569/?lng=FR&supNumDos=FA041553>.
- AFNOR. 2003. *Norme NF EN 13725*. <http://sagaweb.afnor.org/fr-FR/sw/Consultation/Pdf/1267344/?lng=FR&supNumDos=FA100673>.
- AFNOR. 2012. *Norme NF ISO 1600-28 : Air intérieur : Détermination des émissions d'odeurs des produits de construction au moyen de chambre d'essai*.
- AIZPURU, Aitor. 2001. « Biofiltration d'un mélange de COV ». Alès: Université de Pau et des pays de l'Adour.
- AMANO, Saeko, et Hiroyuki UENO. 2004. « Comparative Study of Two Methods for Olfactory Measurement-Triangle Odor Bag method and Dynamic Olfactometry ». <https://doi.org/10.2175/193864704784327304>.
- AMMARI, Faten, Delphine BOUVERESSE JOUAN-RIMBAUD, et Douglas N. RUTLEDGE. 2012. « Two novel methods for the determination of the number of components in Independent Components Analysis models ». http://www.agroparistech.fr/IMG/pdf/Poster_FA_DB_DNR_Afrodata_final.pdf.
- amwalker8. 2018. « Validating Cluster Tendency Using Hopkins Statistic ». *StackExchange*. <https://stats.stackexchange.com/questions/332651/validating-cluster-tendency-using-hopkins-statistic>.
- ANKER, Lawrence S., Peter C. JURIS, et Paul A. EDWARDS. 1990. « Quantitative structure-retention relationship studies of Odor-active aliphatic compounds with oxygen-containing functional group ». *ACS Publications* 62 (24): 2676–2684. <https://doi.org/10.1021/ac00223a006>.
- ANON. 1980. « Reports of Studies on the Measurement of the Offensive Odors (from 1972-1980 in Japan) ». Rapport technique 3588. Health & Environmental Research Online (HERO). Tokyo: Japan Environment Agency.
https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/3588.
- ARLOT, Sylvain. 2017. « Validation croisée ». <https://arxiv.org/pdf/1703.03167.pdf>.
- ASTM. 1997. *Odor Thresholds for Chemicals with Established Occupational Health Standards*. https://books.google.fr/books?id=Lh8nLs9ZuKwC&pg=PA90&lpg=PA90&dq=Ruiken+bij+volledige+vermoeienis,+resp.+adaptatie+voor+een+bepaalden+geur&source=bl&ots=FY11V3Gcvt&sig=UcY5etWcwbq8ElgpaZz-h_KU9zk&hl=fr&sa=X&ved=0ahUKEwjFieP5ha_WAhWGuRoKHQBjB3kQ6AEIJzAA#v=onepage

- &q=Ruiken%20bij%20volledige%20vermoeienis%2C%20resp.%20adaptatie%20voor%20een%20bepaalden%20geur&f=false.
- ASTM International. 1994. *ASTM E178-00 Standard practice for dealing with outlying observation*. <https://www.astm.org/search/fullsite-search.html?query=e178-00&resStart=0&resLength=10&>.
- ATMOTERRA. 2014. « Odeurs et qualité de l'air - Evaluation et gestion des odeurs ». Atmoterra. 2014. <http://www.atmoterra.com/Qualite-Air-et-Odeurs.html>.
- AZENCOTT, Chloé-Agathe. 2017. « Mettez en place un cadre de validation croisée ». Cours. OpenClassrooms. 2017. <https://openclassrooms.com/courses/4297211-evaluez-et-ameliorez-les-performances-dun-modele-de-machine-learning/4308241-mettez-en-place-un-cadre-de-validation-croisee>.
- BACHMANN, C.M., et T.F DONATO. 2000. « An Information Theoretic Comparison of Projection Pursuit and Principal Component Features for Classification of Landsat TM Imagery of Central Colorado ». *Taylor&Francis Online* 21 (15): 2927-35. <https://doi.org/10.1080/01431160050121339>.
- BALABAN, Alexandru T. 1985. « Applications of graph theory in chemistry ». *Journal of chemical information and computer sciences* 25: 334-343.
- BALAKRISHNAN, Naviyn Prabhu, Daphne Gardner TAN SU-LYN, Gade Pandu RANGAIAH, Bee Yong MONG, Goh SU-YEN, et Lakshminarayanan SAMAVEDHAM. 2012. « Partial Least Squares (PLS) Model for Prediction of Definitive and Intermediate Treatment Outcomes in Diabetes Ketoacidosis (DKA) Patients ». In . Singapour: The International Federation of Automatic Control. https://ac.els-cdn.com/S1474667016305432/1-s2.0-S1474667016305432-main.pdf?_tid=6192605c-3428-426d-bef0-e1f9d59fa6ca&acdnat=1533804439_3232f3725892191d1a8523e87001553a.
- BASAK, Subhash C., Ramanathan NATARAJAN, W. NOWAK, P. MISZTA, et Jérôme A. KLUN. 2007. « Three Dimensional Structure-Activity Relationships (3D-QSAR) for Insect Repellency of Diastereoisomeric Compounds: A Hierarchical Molecular Overlay Approach ». *SAR and QSAR in Environmental Research* 18 (3-4): 237-50. <https://doi.org/10.1080/10629360701303784>.
- BAYDAR, A., M. PETRILKA, et M.P. SCHOTT. 1993. « Olfactory thresholds for androstenone and Galaxolide : sensitivity, intensity, and specific anosmia » 18: 661-68.
- BENOIT, Odette, et Gérard PANNER. 1982. « Les polluants atmosphériques et les odeurs-nature, mesures, méthodes d'élimination-synthèse bibliographique ». Rapport de recherche LPC 117. Ministère de l'urbanisme et du logement et ministère des transports - Laboratoire central des ponts et chaussées. http://media.lpc.fr/ext/pdf/ifsttar/librarie/fonds_ancien/RR117.pdf.
- BERGER, Ralf G., Friedrich DRAWERT, Hubert KOLLMANSBERGER, Siegfried NITZ, et Bernhard SCHRAUFSTETTER. 1985. « Novel Volatiles in Pineapple Fruit and Their Sensory Properties ». *ACS Publications* 33 (2): 232-35. <https://doi.org/10.1021/jf00062a019>.
- BERINDE, Zoita. 2006. « Vertex-and edge-weighted molecular graphs for amines ». *Revue Roumaine de Chimie* 51 (11): 1131.
- BERINDE, Zoita. 2008. « QSPR modelling of molar volume of alkanes using the ZEP topological index ». *Creative Math. and Inf* 17: 308-312.
- BERINDE, Zoita, et Mădălina BERINDE. 2004. « On a matrix representation of molecular ». *Carpathian Journal of Mathematics* 20 (2): 205-9.
- BERINDE, Zoita, et Marioara BERINDE. 2009a. « Matrix mathematical models used in the representation of molecular structures ». *Scientific Studies and Research* 19 (2). <http://search.proquest.com/openview/3cc5b62b45a3477b9236e653d2881891/1?pq-origsite=gscholar&cbl=2030309>.
- . 2009b. « Using the topological index ZEP in QSPR studies of alcohols ». *Studi universitatis Babeş-Bolyai, Chimia* 4: 340-52.

- BINARD, Carole. 2013. « La régression PLS - Groupe de travail ». Université de Nice Sophia Antipolis. https://math.unice.fr/~binard/presentation_GDT_PLS_seance2.pdf.
- BLAZY, Vincent, Amaury DE GUARDIA, J.C. BENOIST, M. DAUMOIN, F. GUIZIOU, M. LEMASLE, Dominique WOLBERT, et S. BARRINGTON. 2015. « Correlation of chemical composition and odor concentration for emissions from pig slaughterhouse sludge composting and storage ». *Chemical Engineering Journal* 276: 398-409. <https://doi.org/10.1016/j.cej.2015.04.031>.
- BLONDEL, Mathieu, Kazuhiro SEKI, et Kuniaki UEHARA. 2013. « Block Coordinate Descent Algorithms for Large-Scale Sparse Multiclass Classification ». *Machine Learning* 93 (1): 31-52. <https://doi.org/10.1007/s10994-013-5367-2>.
- BOSE, Bernhard E., Isabelle M. GUYON, et Vladimir N. VAPNIK. 1992. « A Training Algorithm for Optimal Margin Classifier ». In , 144-52. Pittsburgh, Pennsylvania, USA: ACM. <https://doi.org/10.1145/130385.130401>.
- BOUTOU, S., et P. CHATONNET. 2007. « Rapid headspace solid-phase microextraction/gas chromatographic/mass spectrometric assay for the quantitative determination of some of the main odorants causing off-flavours in wine ». *Journal of Chromatography A* 1141 (1): 1-9. <https://doi.org/10.1016/j.chroma.2006.11.106>.
- BOUVERESSE JOUAN-RIMBAUD, Delphine, Adolfo MOYA-GONZALEZ, Faten AMMARI, et Douglas N. RUTLEDGE. 2012. « Two novel methods for the determination of the number of components in independent components analysis models ». *Chemometrics and Intelligent Laboratory Systems* 112: 24-32.
- BREIMAN, Léo. 2001. « Random Forests ». *Kluwer Academic Publishers* 45 (1): 5-32. <https://doi.org/DOI:10.1023/A:1010933404324>.
- BRUCKLER, Franka Miriam. 2008. « Graph theory as a method of improving chemistry and mathematics curricula ». In *Interdisciplinary educational research in mathematics and its connections to the arts and sciences*, 117. IAP. https://books.google.fr/books?id=3TdZ5nrNTEc&printsec=frontcover&hl=fr&source=gbs_ge_summary_r&cad=0#v=onepage&q=curricula&f=false.
- BRUNA, Joan, et Stéphane MALLAT. 2013. « Invariant Scattering Convolution Networks » 35 (8): 1872-86.
- BUCK, L., et R. AXEL. 1991. « A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition ». *Cell*. 65 (1): 175-87. [https://doi.org/10.1016/0092-8674\(91\)90418-X](https://doi.org/10.1016/0092-8674(91)90418-X).
- CAIN, William S., et Roland SCHMIDT. 2009. « Can we trust odor databases? Example of t- and n-butyl acetate » 43: 2591-2601. <https://doi.org/doi:10.1016/j.atmosenv.2009.02.024>.
- CANDAU, Joël, et Agnès JEANJEAN. 2006. « Des odeurs à ne pas regarder ». *Odeurs*, n° 47: 51-68. <https://doi.org/10.4000/terrain.4251>.
- CAPELLI, Laura, Selena SIRONI, Renato DEL ROSSO, et Jean-Michel GUILLOT. 2013. « Measuring odours in the environment vs. dispersion modelling: A review ». *Atmospheric Environment* 79: 731-43. <http://dx.doi.org/10.1016/j.atmosenv.2013.07.029>.
- CHAIGNAUD, Mathilde, Stéphane CARIOU, Julie POETTE, Marion FAGES, Jean-françois DESPRES, et Jean-Louis FANLO. 2014. « A new method to evaluate odour annoyance potential ». *AIDIC* 40. <https://doi.org/10.3303/CET1440003>.
- CHAKRAPANI, Chuck. 2011. « Shamanism as Statistical Knowledge: Is a Sample Size of 30 All You Need? », 2011.
- CHAPUT, Brigitte, Christelle FERRER, Jean-Philippe GARCIA, Alain GARDIENNET, Delphine JUGAN, Yannick LE BASTARD, Fabrice LECLERCQ, et al. 2013. « Test de dixon recherche de valeurs aberrantes ». Bulletin 22. PY-MATH. ENFA. <http://r2math.enfa.fr/wp-content/uploads/sites/8/2013/06/22-7-dixon.pdf>.
- CHAVENT, Marie, et Brigitte PATOUILLE. 2003. « Calcul des coefficients de régression et du PRESS en régression PLS1 », n° 30.

- CHENG, Tiejun, Yuan ZHAO, Xun LI, Fu LIN, Yong XU, Xinglong ZHANG, Yan LI, Renxiao WANG, et L. LAI. 2007. « Computation of octanol-water partition coefficients by guiding an additive model with knowledge ». *Journal of Chemical Information and Modeling* 6 (47): 2140-48. <https://doi.org/10.1021/ci700257y>.
- CHONG, Il-Gyo, et Chi-Hyuck JUN. 2005. « Performance of some variable selection methods when multicollinearity is present ». *Elsevier* 78 (1-2): 103-12. <https://doi.org/10.1016/j.chemolab.2004.12.011>.
- CHOUAF, Seloua, et Youcef SMARA. 2012. « Méthode de sélection des bandes à base de l'analyse en composantes indépendantes appliquée aux images hyperspectrales de télédétection ». *Revue Française de Photogrammétrie et de Télédétection*, n° 204: 57-62.
- COMETTO-MUNIZ, J. Enrique, et William S. CAIN. 1994. « Sensory reactions of nasal pungency and odor to volatile organic compounds : the alkylbenzenes » 55 (9): 811-17.
- COMETTO-MUNIZ, J. Enrique, William S. CAIN, et Michael H. ABRAHAM. 2004. « Detection of singles and mixed VOCs by smell and by sensory irritation » 14 (Suppl. 8): 108-17.
- CONSONNI, Viviana, Roberto TODESCHINI, Manuela PAVAN, et Paola GRAMATICA. 2002. « Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies ». *Journal of Chemical Information and Computer Sciences* 42 (3): 693-705. <https://doi.org/10.1021/ci0155053>.
- CORDERO, C., J. KIEFL, P. SCHIEBERLE, S.E. REICHENBACH, et C. BICCHI. 2015. « Comprehensive two-dimensional gas chromatography and food sensory properties: Potential and challenges ». *Analytical and Bioanalytical Chemistry* 407 (1): 169-91. <https://doi.org/10.1007/s00216-014-8248-z>.
- CORNUEJOLS, Antoine*. 2002. « Une nouvelle méthode d'apprentissage : Les SVM. Séparateurs à vaste marge ». *Bulletin de l'AFIA*, n° 51. <https://www.lri.fr/~antoine/Papers/SVM-synthese.pdf>.
- CRAMER III, Richard D. 1993. « Partial Least Squares (PLS): Its strengths and limitations ». *Springer* 1 (2): 269-78.
- CZERNY, M., R. BRUECKNER, E. KIRCHHOFF, R. SCHMITT, et A. BUETTNER. 2011. « The Influence of Molecular Structure on Odor Qualities and Odor Detection Thresholds of Volatile Alkylated Phenols ». *Chemical Senses* 36 (6): 539-53. <https://doi.org/10.1093/chemse/bjr009>.
- D'ACAMPORA ZELLNER, Barbara, Paola DUGO, Giovanni DUGO, et Luigi MONDELLO. 2008. « Gas chromatography-olfactometry in food flavour analysis ». *Journal of Chromatography A* 1186 (1-2): 123-43. <https://doi.org/10.1016/j.chroma.2007.09.006>.
- DAVID, Sophie, Danièle DUBOIS, Catherine ROUBY, et Benoist SCHAAL. 1997. « L'expression des odeurs en français : Analyse lexicale et représentation cognitive ». *Intellectica* 1 (24): 51-83.
- DAVOLI, Enrico, L. GANGAI, L. MORSELLI, et D. TONELLI. 2003. « Characterisation of odorants emissions from landfills by SPME and GC/MS ». *Chemosphere* 51 (5): 357-68. [https://doi.org/10.1016/S0045-6535\(02\)00845-7](https://doi.org/10.1016/S0045-6535(02)00845-7).
- DE MAESSCHALCK, R., F. ESTIENNE, J. VERDU-ANDRES, A. CANDOLFI, V. CENTNER, F. DESPAGNE, D. JOUAN-RIMBAUD, et al. 1999. « The_development_of_calibration_models_for_spectros.pdf ». *Internet Journal of Chemistry* 2 (19).
- DE MARCH, Claire. 2015. « Modélisation des mécanismes moléculaires de la perception des odeurs ». Université Nice Sophia Antipolis. <https://tel.archives-ouvertes.fr/tel-01432834/document>. <https://tel.archives-ouvertes.fr/tel-01432834>.
- DEAN, R. B., et W. J. DIXON. 1951. « Simplified Statistics for Small Numbers of Observations ». *Analytical Chemistry* 23 (4): 636-38. <https://doi.org/10.1021/ac60052a025>.

- DECOTTIGNIES, V., A. HUYARD, R.F. KELLY, et B. BARILLON. 2013. « Development of a diagnostic tool: the wastewater collection network odour wheel ». *Water Science and Technology* 68 (4): 839-47. <https://doi.org/10.2166/wst.2013.265>.
- DECREMER, Renaud. 2006. « La régression PLS ». Mémoire. Université paul verlaine de METZ. http://www.memoireonline.com/05/08/1083/m_la-regression-pls2.html.
- DELVA, Julien, Pierre COBUT, et Jean-Louis FANLO. 2017. « De l'odeur à l'impact sur le territoire ». *Pollution atmosphérique*, n° 234: 13. <https://doi.org/10.4267/pollution-atmospherique.5232>.
- DEMOLE, Edouard, et DR. M. STOLL. 1964. « Synthèse et relations entre constitution chimique et odeur dans la série des terpényl-3-cyclohexanols ». *Helvetica chimica acta* 47 (7): 1766-74. <https://doi.org/10.1002/hlca.19640470713>.
- DESAUZIERS, Valérie. 2009. « Analyse des COV en traces dans l'air ». *Les techniques de l'ingénieur* G1634 v1: 1-24.
- DEVOS, M., F. PATTE, J. ROUAULT, L.J. LAFFORT, et L.J. VAN GEMERT. 1990. *Standardized human olfactory thresholds*. Oirl Press.
- DISTEL, H., S. AYABE-KANAMURA, M. MARTINEZ-GOMEZ, I. SCHICKER, T. KOBAYAKAWA, S. SAITO, et R. HUDSON. 1999. « Perception of everyday odors: Correlation between intensity, familiarity and strength of hedonic judgement ». *Chemical Senses* 24 (2): 191-99.
- DIUDEA, M.V., Ovidiu IVANCIUC, Sonja NIKOLIC, et Nenad TRINAJSTIC. 1997. « Matrices of Reciprocal Distance ». *MATCH Communications in mathematical and in computer chemistry*, n° 35.
- DOBRYNIN, Andrey A. 1999. « A simple formula for the calculation of of the wiener index of hexagonal chains ». *Computers & chemistry* 23 (1): 43-48.
- DON, J.A. 1986. « Odour measurement and control, paper presented at "Odour control in industry" ». In , 18:1-18. Londres, Angleterre.
- DRAVNIKS, Andrew. 1974. « A Building-Block Model for the Characterization of Odorant Molecules and Their Odors ». *Annals of the New York Academy of Sciences* 237 (1): 144-163. <https://doi.org/10.1111/j.1749-6632.1974.tb49851.x>.
- DRAVNIKS, Andrew, et William H. PROKOP. 1975. « Source Emission Odor Measurement by a Dynamic Forced-Choice Triangle Olfactometer ». *Journal of the Air Pollution Control Association* 25 (1): 28-35. <https://doi.org/10.1080/00022470.1975.10470045>.
- DUONG, Tarn. 2001. « An introduction to kernel density estimation ». <http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/>.
- DUROCHER, Martin. 2015. « Approches statistiques pour la prédiction de variables hydrologiques extrêmes à des sites non jaugés en l'absence d'hypothèses de linéarité et normalité ». Québec: Université du Québec. <http://espace.inrs.ca/3340/>.
- DUROCHER, Martin, Fateh CHEBANA, et Taha B. M. J. OUARDA. 2015. « A Nonlinear Approach to Regional Flood Frequency Analysis Using Projection Pursuit Regression ». *Journal of Hydrometeorology* 16 (4): 1561-74. <https://doi.org/10.1175/JHM-D-14-0227.1>.
- EDUSCOL. 2016. « Recherche de valeurs aberrantes Test de Dixon ». Portail. Ressources Nationales de Chimie. 2016. <http://eduscol.education.fr/rnchimie/math/benichou/tests/dixon/dixon.htm>.
- EduSoft LC. s. d. « Methods background topological indices.pdf ». In *Molconn-Z 4.00*. <http://www.edusoft-lc.com/molconn/manuals/400/chaptwo.html>.
- EPA - Environmental Protection Agency. 2001. « Odour Impacts and Odour Emission Control Measures for Intensive Agriculture ». R&D report - Environmental Research 14. 1-84095-075-7. <http://www.epa.ie/pubs/reports/research/air/Odour%20Impacts%20Final.pdf>.
- EYKMAN, Hendriks Johannes. 1927. *Klinische olfactometrie*. Bibliotheek der Rijkuniversiteit Utrecht. Utrecht, Pays-Bas. <http://objects.library.uu.nl/reader/index.php?obj=1874-287120&lan=en#page//89/67/05/89670508452178594110146005895544190608.jpg/mode/1up>.

- FALCY, Michel, et Stéphane MALARD. 2005. « Comparaison des seuils olfactifs de substances chimiques avec des indicateurs de sécurité utilisés en milieu professionnel ». *Hygiène et Sécurité au travail- Cahier de notes documentaire (INRS)*, n° ND2221-198-05: 7-21.
- FERNANDEZ, Xavier, Katharina BREME, et Vincent VARLET. 2009. « Couplage chromatographie en phase gazeuse/olfactométrie ». *Techniques de l'ingénieur. Analyse et caractérisation 1 (P1488)*: 1-5.
- FLIGHT, Robert M. 2015. « Random Forest vs PLS on Random Data ». Statistique. Deciphering Life : One Bit at a Time. 2015. <https://rmflight.github.io/post/random-forest-vs-pls-on-random-data/>.
- FORS, Susan M., et Bertil K. OLOFSSON. 1985. « Alkylpyrazines, volatiles formed in the Maillard reaction. I. Determination of odour detection thresholds and odour intensity functions by dynamic olfactometry » 10 (3): 287-96. <https://doi.org/10.1093/chemse/10.3.287>.
- FRANCOIS, Jean-Christophe, et Claude GRASLAND. 2000. « Chapitre STAT.4 : Les paramètres de dispersion (GO152 : Introduction à la statistique et à la cartographie en géographie) ». <http://grasland.script.univ-paris-diderot.fr/Go152/Stat5/stat5.htm>.
- FRIEDMAN, Jerome H. 1985. « Exploratory Projection Pursuit ». *Journal of the American Statistical Association* 82 (397): 249-66.
- FRIEDMAN, Jerome H., et John W. TUKEY. 1974. « A Projection Pursuit Algorithm for Exploratory Data Analysis ». *IEEE Transactions on Computers* C-23 (9): 881-90. <https://doi.org/10.1109/T-C.1974.224051>.
- GALLEGO, E., F.J. ROCA, J.F. PERALES, G. SANCHEZ, et P. ESPLUGAS. 2012. « Characterization and Determination of the Odorous Charge in the Indoor Air of a Waste Treatment Facility through the Evaluation of Volatile Organic Compounds (VOCs) Using TD-GC/MS ». *Waste Management* 32 (12): 2469-81. <https://doi.org/10.1016/j.wasman.2012.07.010>.
- GANTY, Philippe. 2016. « L'art de la dégustation ». Rhumagricole.ch - Terroir et Authenticité. 2016. <https://rhumagricole.ch/lart-de-la-degustation/>.
- Gemert, L J Van. 2011. *Odour thresholds. Compilations of odour threshold values in air, water and other media. (second enlarged and revised edition)*. Utrecht (The Netherlands): Oliemans Punter & Partners BV.
- GHARAGHEIZI, Farhad, Ali ESLAMIMANESH, Poorandokht ILANI-KASHKOULI, Amir H. MOHAMMADI, et Dominique RICHON. 2012. « QSPR Molecular Approach for Representation/Prediction of Very Large Vapor Pressure Dataset ». *Chemical Engineering Science* 76 (juillet): 99-107. <https://doi.org/10.1016/j.ces.2012.03.033>.
- GILBERT, Laura. 2012. « Caractérisation physico-chimique et sensorielle d'ingrédients cosmétiques : une approche méthodologique ». Le Havre: Université du Havre. https://tel.archives-ouvertes.fr/file/index/docid/956601/filename/MANUSCRIT_FINALISE_-_25_juin_2013.pdf.
- GOLDFARB, Bernard, et Catherine PARDOUX. 2005. « Méthodes d'ajustements graphiques : Diagramme Quantile – Quantile ». *MODULAD*, n° 33: 1-6.
- GONZALEZ, Pierre-Louis, et Michel TENENHAUS. 2000. « Les méthodes PLS ». http://maths.cnam.fr/IMG/pdf/REGRESSION_PLS-1-2-Disc-Logist.pdf.
- GOULON-SIGWALT-ABRAM, Aurélie. 2008. « Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments ». Université Pierre et Marie Curie-Paris VI.
- GRASLAND, Claude. 2000. « Chapitre 5 : Transformation et comparaison de caractères quantitatifs ». http://grasland.script.univ-paris-diderot.fr/STAT98/stat98_5/stat98_5.htm.
- GRIMM, Casey, Elaine CHAMPAGNE, et Ken-Ichi OHTSUBO. 2001. « Analysis of volatile compounds in the headspace of rice using SPME/GC/MS ». In *Flavor, Fragrance, and Odor Analysis*, CRC Press, 1:229-49. Boca Raton: MARSILI R. https://www.researchgate.net/publication/290143615_Analysis_of_Volatile_Compounds_in_the_Headspace_of_Rice_Using_SPME_GC_MS.

- GUIGOURES, Romain. 2013. « Utilisation des modèles de co-clustering pour l'analyse exploratoire des données ». Paris: Université Panthéon-Sorbonne-Paris I. <https://tel.archives-ouvertes.fr/tel-00935278/document>.
- GUTIERREZ, M.C., A.F. CHICA, M.A. MARTIN, et A.C. ROMAIN. 2014. « Compost pile monitoring using different approaches: GC-MS, E-nose and dynamic olfactometry ». *Waste and Biomass Valorization* 5 (3): 469-79. <https://doi.org/10.1007/s12649-013-9240-0>.
- GUTMAN, Ivan. 1994. « selected properties of the schultz molecular topological index ». *ACS Publications* 5 (34): 1087-89. <https://doi.org/10.1021/ci00021a009>.
- GUTMAN, Ivan, et Sandi KLAUZAR. 1997. « Bounds for the Schultz molecular topological index of benzenoid systems in terms of Wiener index ». *ACS* 4 (37): 741-44. <https://doi.org/10.1021/ci9700034>.
- GUTMAN, Ivan, et Nenad TRINAJSTIC. 1972. « Graph Theory and Molecular Orbitals.Total Pi-Electron Energy of Alternant Hydrocarbons ». *Chemical Physics Letters* 17 (4): 535-38.
- HALL, Lowell H., et Lemont B. KIER. 1995. « Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information ». *Journal of Chemical Information and Modeling* 35 (6): 1039-45. <https://doi.org/10.1021/ci00028a014>.
- HALL, Lowell H., Brian MOHNEY, et Lemont B. KIER. 1991. « The electrotopological state: an atom index for QSAR ». *Molecular Informatics* 10 (1): 43-51. <https://doi.org/10.1002/qsar.19910100108>.
- HAN, Jiawei, Micheline KAMBER, et Jian PEI. 2012. *Data Mining Concepts and Techniques, Third Edition*.
file:///C:/Users/Pascale/AppData/Local/Microsoft/Windows/INetCache/IE/BBV966TF/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf.
- HAN, K.H., J.S. ZHANG, H.N. KNUDSEN, P. WARGOCKI, H. CHEN, P.K. VARSHNEY, et B. GUO. 2011. « Development of a novel methodology for indoor emission source identification ». *Atmospheric Environment* 45 (18): 3034-45. <https://doi.org/10.1016/j.atmosenv.2011.03.021>.
- HANSCH, Corwin, et AJ LEO. 1979. *Substituent constants for correlation analysis in chemistry and biology*. John Wiley and Sons. New York: John Wiley and Sons.
- HELLMAN, Thomas M., et Francis H. SMALL. 1974. « Characterization of the Odor Properties of 101 Petrochemicals Using Sensory Methods ». *Journal of the Air Pollution Control Association* 24 (10): 979-82. <https://doi.org/10.1080/00022470.1974.10470005>.
- HEMMATEENEJAD, Bahram, Ramin MIRI, et Maryam ELYASI. 2012. « A Segmented Principal Component Analysis—Regression Approach to QSAR Study of Peptides ». *Journal of Theoretical Biology* 305 (juillet): 37-44. <https://doi.org/10.1016/j.jtbi.2012.03.028>.
- HERVÉ, Maxime. 2014. « Aide-mémoire de statistique appliquée à la biologie ». *Constr Son Étude Anal Résultats à l'aide du logiciel R. Version 5*.
<http://mirror.psu.ac.th/pub/cran/doc/contrib/Herve-Aide-memoire-statistique.pdf>.
- HIGUCHI, Takaya. 2004. « Odor measurement in Japan ». In *East Asia workshop on odor measurement and control review, Practical problems of odor measurements and odor controls*. Available from: http://www.env.go.jp/en/air/odor/eastasia_ws/index.htm. https://chubu.env.go.jp/en/air/odor/eastasia_ws/2-1-5.pdf.
- HOGG, Robert V., et Elliot A. TANIS. 2005. *Probability and statistical inference (7th Edition)*. Prentice Hall. New Jersey: Pearson.
- HOPKINS, Brian, et John Gordon SKELLAM. 1954. « A New Method for determining the Type of Distribution of Plant Individuals » 18 (2): 213-27.
- HUA, Hongbo. 2009. « Wiener and Schultz molecular topological indices of graphs with specified cut edges ». *MATCH Communications in mathematical and in computer chemistry* 61 (3): 643.
- HUBER, Peter J. 1985. « Projection Pursuit ». *The Annals of Statistics* 13 (2): 435-75.

- HUIJER, Cornelis. 1917. *De olfactologie van aniline en homologen*. Utrecht, Pays-Bas: A. Oosthoek.
<https://catalog.hathitrust.org/Record/005676128>.
- HYVARINEN, Aapo. 2010. « Independent Component Analysis and Blind Source Separation ».
 présenté à InterBrain Symposium, Jyväskylä.
<http://users.jyu.fi/~wijaje/interbrain/presentations/Hyvarinen.pdf>.
- HYVARINEN, Aapo, et Erkki OJA. 2000. « Independent Component Analysis: Algorithms and Applications ». *Neural Networks* 13 (4-5): 411-30. [https://doi.org/10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5).
- IBM. 2012. « Oracle Support Vector Machine (SVM) ». IBM Knowledge Center. 2012.
https://www.ibm.com/support/knowledgecenter/fr/SS3RA7_15.0.0/com.ibm.spss.modeler.help/oracle_svm.htm.
- INRS. 2000. « Le calcul d'incertitude dans les méthodes de mesurage de l'exposition professionnelle ». *Métropol - Fiche G*, 1-20.
- ISO17025. 2015. « Tests de comparaison des valeurs mesurées ». DemarcheISO17025. 2015.
<http://www.demarcheiso17025.com/fiche016.html>.
- IVANCIUC, Ovidiu. 1998. « Canonical numbering and constitutional symmetry ». In *The encyclopedia of computational chemistry John Wiley & Sons*, 167-83. Chichester.
- IVANCIUC, Ovidiu, Teodor-Silviu BALABAN, et Alexandru T. BALABAN. 1993. « Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices ». *Journal of Mathematical Chemistry* 12 (1): 309-318.
- JAUBERT, Jean-Noël, G. GORDON, et J.C. DORE. 1987a. « Une organisation du champ des odeurs I ». *Parfums, cosmétiques, arômes*, n° 77: 53-56.
- . 1987b. « Une organisation du champ des odeurs II ». *Parfums, cosmétiques, arômes*, n° 78: 71-82.
- JAVEAU, Alain. 2016. « L'information, en tant que lien essentiel entre les sciences de la nature et les sciences humaines ». Dijon: Université de Bourgogne-Franche-Comté.
 File:///C:/Users/Pascale/Zotero/storage/H86XRWBU/2016-A_JAVEAU_Alain_th%C3%A8seentropie.pdf. <https://tel.archives-ouvertes.fr/tel-01529752/document>. <https://tel.archives-ouvertes.fr/tel-01529752>.
- JOLLIFFE, I.T. 2002. *Principal Component Analysis, Second Edition*. Springer. New-York.
[http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf).
- JONES, M.C., et Robin SIBSON. 1987. « What Is Projection Pursuit? » *Journal of the Royal Statistical Society. Series A* 150 (1): 1-37. <https://doi.org/10.2307/2981662>.
- JUHLKE, Florian, Katja LORBER, Maria WAGENSTALLER, et Andrea BUETTNER. 2017. « Influence of the Chemical Structure on Odor Qualities and Odor Thresholds of Halogenated Guaiacol-Derived Odorants ». *PubMed* 5 (120). <https://doi.org/10.3389/fchem.2017.00120>.
- JUNG, J. 1936. « Untersuchung über Geruchsschwellen ». Würzburg: Code E2.
- KAR, Sitanshu Sekhar, et Archana RAMALINGAM. 2013. « Is 30 the Magic Number? Issues in Sample Size Estimation ». *Open Access Article* 4 (1): 175-79.
- kassambra. 2017. « Assessing Clustering Tendency : Essentials ». Statistiques. Statistical Tools for high-throughput data analysis. 2017. <http://www.sthda.com/english/articles/29-cluster-validation-essentials/95-assessing-clustering-tendency-essentials/>.
- KERMEN, F., A. CHAKIRIAN, C. SEZILLE, P. JOUSSAIN, G. LE GOFF, A. ZIESSEL, M. CHASTRETTE, et al. 2011. « Molecular Complexity Determines the Number of Olfactory Notes and the Pleasantness of Smells ». *Scientific Reports* 1 (1). <https://doi.org/10.1038/srep00206>.
- KIER, Lemont B., et Lowell H. HALL. 1990. « An electrotopological-state index for atoms in molecules » 7 (8): 801-7.
- KnowledgeDoor LLC. 2017. « Sanderson Electronegativity ». KnowledgeDoor. 2017.
http://www.knowledgedoor.com/2/elements_handbook/sanderson_electronegativity.html.

- KOGA, Minoru, Yoshifumi HANADA, Junlin ZHU, et Osamu NAGAFUCHI. 2001. « Determination of ppt levels of atmospheric volatile organic compounds in Yakushima, a remote south-west island of Japan ». *Microchemical Journal* 68 (2-3): 257-64. [https://doi.org/10.1016/S0026-265X\(00\)00151-X](https://doi.org/10.1016/S0026-265X(00)00151-X).
- KOSTER, Egon Peter. 1971. *Adaptation and cross-adaptation in olfaction : An Experimental Study with Olfactory Stimuli at Low Levels of Intensity*. Brönder-Offset.
- KOSTER, E.P. 1991. « Méthodes d'évaluation psychologiques dans l'étude de l'environnement – In : Martin G., Laffort P. (dir.), Odeurs & Désodorisation, Dans l'environnement ». In *Odeurs & Désodorisation dans l'environnement*, Tec&Doc Lavoisier, 25-58. Paris: MARTIN G., LAFFORT P.
- KRAFT, P., et K. POPAJ. 2004. « Total synthesis and olfactory evaluation of 5beta,10-dimethyl-des-A-18-nor-androstan-13beta-ol: a potential human pheromone? » 24: 4995–5002.
- LARMARANGE, Joseph. 2018. « Classification ascendante hiérarchique ». GitHub. 2018. <http://larmarange.github.io/analyse-R/classification-ascendante-hierarchique.html>.
- LAROUSSE. 2018. « Néguentropie ». Dictionnaire en ligne. Larousse. 2018. <https://www.larousse.fr/dictionnaires/francais/n%C3%A9gumentropie/54099>.
- LARSEN, Dr. David R. 2014. « Hopkins' Index of Aggregation ». Natural Resource Biometrics. 2014. <http://oak.snrmissouri.edu/nr3110/topics/hopkins.php>.
- LASKA, Matthias. 2010. « Olfactory Perception of 6 Amino Acids by Human Subjects » 35 (4): 279-87. <https://doi.org/10.1093/chemse/bjq017>.
- LAWSON, Richard G., et Peter C. JURIS. 1990. « New Index for Clustering Tendency and Its Application to Chemical Problems ». *ACS Publications* 30 (1): 36-41. <https://doi.org/10.1021/ci00065a010>.
- LEBARBIER, E, et T MARY-HUARD. 2008. « Classification non supervisée ». Polycopié de cours AgroParisTech. <https://tice.agroparistech.fr/coursenligne/courses/STATISTIQUESCD4/document/TP-Classif/ClassificationNonSupervisee.pdf?cidReq=STATISTIQUESCD4>.
- LEE, SY, A. MEDIANI, M. MAULIDIANI, A. KHATIB, IS ISMAIL, N. ZAWAWI, et F. ABAS. 2018. « Comparison of partial least squares and random forests for evaluating relationship between phenolics and bioactivities of *Neptunia oleracea* » 98 (1). <https://doi.org/10.1002/jsfa.8462>.
- LENNON, M., G. MERCIER, M.C. MOUCHOT, et Laurence HUBERT-MOY. 2001. « Independent Component Analysis as a Tool for the Dimensionality Reduction and the Representation of Hyperspectral Images ». In , 6:2893-95. Sydney: IEEE. <https://doi.org/10.1109/IGARSS.2001.978197>.
- LEONARDOS, Gregory, David KENDALL, et Nancy BARNARD. 1969. « Odor Threshold Determinations of 53 Odorant Chemicals ». *Journal of the Air Pollution Control Association* 19 (2): 91-95. <https://doi.org/10.1080/00022470.1969.10466465>.
- LESTREMAU, François. 2003. « Développement d'une méthodologie analytique pour le dosage de composés soufrés malodorants dans différentes matrices environnementales ». Pau: École doctorale des sciences exactes et des leurs applications.
- LI, P.W., Z.H. YU, et H. LI. 1992. *Food Flavor and Fragrance Chemistry: Heterocyclic aroma Chemicals*. China: China Light Industry Press.
- LINDELL, Henry. 1991. *A study of odour and taste originating from food packaging board analysed by chromatographic techniques and sensory evaluation*. Åbo Academy Press. Vol. 51. 4 vol. Acta Academiae Aboensis Mathematica et physica. Abo.
- LINDINGER, Christian, David LABBE, Philippe POLLIEN, Andreas RYTZ, Marcel A. JUILLERAT, Chahan YERETZIAN, et Andlme BLANK. 2008. « When Machine Tastes Coffee: Instrumental Approach To Predict the Sensory Profile of Espresso Coffee » 80 (5): 1574-81.
- LLEDO, Pierre-Marie, et Jean-Didier VINCENT. 1999. « Odeurs » 15 (11): 1211-18.

- LO, Y.C., J.A. KOZIEL, L. CAI, S.J. HOFF, W.S. JENKS, et H. XIN. 2008. « Simultaneous Chemical and Sensory Characterization of Volatile Organic Compounds and Semi-Volatile Organic Compounds Emitted from Swine Manure using Solid Phase Microextraction and Multidimensional Gas Chromatography–Mass Spectrometry–Olfactometry ». *Journal of Environment Quality* 37 (2): 521–34. <https://doi.org/10.2134/jeq2006.0382>.
- LOHNINGER, H. 2012. « Predictive Ability ». *Statistiques. Fundamentals of Statistics*. 2012. http://www.statistics4u.com/fundstat_eng/cc_press.html.
- MA, Da, Apurba K. BHATTACHARJEE, Raj K. GUPTA, et Jean M. KARLE. 1999. « Predicting mosquito repellent potency of N, N-diethyl-m-toluamide (DEET) analogs from molecular electronic properties. » *The American journal of tropical medicine and hygiene* 60 (1): 1–6.
- MAARSE, H., L.M. NIJSSEN, et S.A.G.F. ENGELINO. 1988. « Halogenated Phenols and Chloroanisoles: Occurrence, Formation and Prevention ». In *Characterization, Production and Application of Food Flavours*, Akademik Verlag, 43–63. Berlin, Belgique: Rothe M. Ed.
- MAILLE, Virginie. 2000. « La réponse du consommateur au produit parfumé : l'incidence de la congruence de l'odeur ». In , 1–28. ESCP Europe / Ca' Foscari University, Venise, Italie. <http://archives.marketing-trends-congress.com/2000/pdf/maille.pdf>.
- MALDONADO, Ana. 2006. « Diversité Moléculaire: Application au Criblage Virtuel, Corrélation avec des Propriétés Physico-chimiques ». Université Paris 7. http://www.agmaldonado.com/perso/PhD_AnaMaldonado.pdf.
- MALLAT, Stéphane. 2015. « Learning Multiscale Invariants from Big Data for Physics ». ENS Paris. <https://www.phys.ens.fr/spip.php?article2207>.
- MARCHINI, J.L., C. HEATON, et B.D. RIPLEY. 2017. « Package “FastICA” ». CRAN. <https://cran.r-project.org/web/packages/fastICA/fastICA.pdf>.
- MAY, J. 1966. « Geruchsschwellen von Lösemitteln zur Bewertung von Lösemittelgerüchen in der Luft (Odor thresholds of solvents for the assessment of solvent odors in the air) », n° 26: 385–89.
- MEDJKOUNE, Massissilia, Sébastien HARISPE, Jacky MONTMAIN, Stéphane CARIOU, Jean-Louis FANLO, et Nicolas FIORINI. 2016. « Towards a non-oriented approach for the evaluation of Odor quality ». In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 610:238–49. Springer. https://doi.org/10.1007/978-3-319-40596-4_21.
- MEIERHENRICH, Uwe J., Jérôme GOLEBIEWSKI, Xavier FERNANDEZ, et Daniel CABROL-BASS. 2005. « De la molécule à l'odeur ». *L'actualité chimique* 289: 29.
- MILAN, Bianca J.B., Simon S.K. BOOTSMA, et Ilse BILSEN. 2012. « Advances in odour monitoring with e-noses in the port of rotterdam ». *AIDIC* 30: 145–50. <https://doi.org/10.3303/CET1230025>.
- . 2016. « eNoses as a Tool to Measure Odour Nuisance Caused by Restaurants ». *AIDIC* 54: 61–66. <https://doi.org/10.3303/CET1654011>.
- MILLER, James N. 2015. « Using the Grubbs and Cochran Tests to Identify Outliers ». *The Royal Society of Chemistry, AMC Technical Briefs*, 19 (69): 1–3. <https://doi.org/10.1039/c5ay90053k>.
- Minitab 18. 2017. « Interprétation de toutes les statistiques et de tous les graphiques pour la fonction Test t à 2 échantillons ». Support. Assistance de Minitab 18. 2017. <https://support.minitab.com/fr-fr/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/2-sample-t/interpret-the-results/all-statistics-and-graphs/>.
- MOLHAVE, L., S.K. KJAERGAARD, A. HEMPELJORGENSEN, J.E JUTO, K. ANDERSSON, G. STRIDH, et J. FALK. 2000. « The Eye Irritation and Odor Potencies of Four Terpenes Which Are Major Constituents of the Emissions of VOCs from Nordic Soft Woods ». *PubMed* 10 (4): 315–18.
- MONTESTRUCQ, Luciles, Olivier GUYE, Elsa PEYSSON, et Hadrien CHARVAT. 2007. « Tableau de bord Santé Environnement Rhône-Alpes : La qualité de vie, le bruit et les odeurs ». Tableau de bord. http://www.ors-rhone-alpes.org/environnement/pdf/Contaminants_nuisances.pdf. http://www.ors-auvergne-rhone-alpes.org/environnement/pdf/Contaminants_nuisances.pdf.

- MORICE, E. 1972. « Tests de normalité d'une distribution observée ». *NUMDAM* 20 (2): 5-35.
- MORIWAKI, Hiroto, Yu-Shi TIAN, et Tatsuya TAKAGI. 2018. « Mordred: a molecular descriptor calculator ». *Springer* 10 (4). <https://doi.org/10.1186/s13321-018-0258-y>.
- MORRISON, E.E., et R.M. COSTANZO. 1990. « Morphology of the human olfactory epithelium » 297 (1): 1-13.
- MOZRZYMAS, Anna. 2011. « A Generalized Graph-Theoretical Matrix of Heterosystems and Its Application to the VMV Procedure ». *Computational Biology and Chemistry* 35 (6): 347-52. <https://doi.org/10.1016/j.compbiolchem.2011.10.004>.
- MU, Lailong, Changjun FENG, et Lu XU. 2006. « Study on QSPR of Alcohols with a Novel Edge Connectivity Index mF ». *MATCH Communications in mathematical and in computer chemistry* 56: 217-230.
- MUNOZ, Raul, Eric C. SIVRET, Gavin PARCSI, LEBRERO R., Xinguang WANG, I.H. SUFFET, et Richard M. STUETZ. 2010. « Monitoring techniques for odour abatement assessment » 44 (18): 5129-49. <https://doi.org/10.1016/j.watres.2010.06.013>.
- NAGATA, Yoshio. 2003. « Measurement of Odor Threshold by Triangle Odor Bag Method ». *Odor Measurement Review, Office of Odor, Noise and Vibration Environmental Management Bureau, Ministry of the Environment, Government of Japan, Tokyo, Japan*, 118-127.
- NAGY, George Z. 1991. « The Odor Impact Model ». *Journal of the Air & Waste Management Association* 41 (10): 1360-62. <https://doi.org/10.1080/10473289.1991.10466934>.
- NATARAJAN, Ramanathan, Subhash C. BASAK, Alexandru T. BALABAN, Jérôme A. KLUN, et Walter F. SCHMIDT. 2005. « Chiral index, molecular overlay and biological activity of diastereoisomeric mosquito repellent ». *Pest Management Sciences*, n° 61: 1193-1201. <https://doi.org/10.1002/ps.1116>.
- NAUS, Antonin. 1982. « Olfactory thresholds of some industrial substances » 34: 217-18.
- NESTEROV, Igor V., Andrey A. TOROPOV, Pablo R. DUCHOWICZ, et Eduardo A. CASTRO. 2004. « An Improved QSPR Modeling of Hydrocarbon Dipole Moments ». *The Scientific World JOURNAL* 4: 956-64. <https://doi.org/10.1100/tsw.2004.193>.
- NICELL, Jim A. 2009. « Assessment and Regulation of Odour Impacts ». *Atmospheric Environment* 43 (1): 196-206. <https://doi.org/10.1016/j.atmosenv.2008.09.033>.
- NICOLAS, Jacques, Anne-Claude ROMAIN, et Julien DELVA. 2008. « Les odeurs dans l'environnement : dimensions sensorielles et méthodes d'évaluation ». *Journal des ingénieurs*, n° 109: 12-15.
- NICOLLE, Jérôme. 2009. « Développement d'une méthodologie d'analyse de composés organiques volatils en traces pour la qualification de matériaux de construction ». Pau: Université de Pau et des pays de l'Adour. <https://tel.archives-ouvertes.fr/tel-00401405/document>.
- NIH. 2004. « PubChem-Open Chemistry Database ». PubChem-Open Chemistry Database. 2004. <https://pubchem.ncbi.nlm.nih.gov/compound/90728#section=Identification>.
- NIKOLIC, Sonja, Nenad TRINAJSTIC, et Zlatko MIHALIC. 1993. « Molecular topological index: An extension to heterosystems ». *Journal of mathematical chemistry* 12 (1): 251-264.
- NIST. 2017. « WebBook de chimie NIST (National Institute of Standards and Technology) ». Banque de données. NIST WebBook de chimie NIST, SRD 69. 2017. <http://webbook.nist.gov/chemistry/>.
- OBOZINSKI, Guillaume. 2013. « Apprentissage: cours 3, Validation croisée, Consistance uniforme, Théorème No Free Lunch ». http://www.math.ens.fr/cours-apprentissage/Obozinski/cours_2013/cours3.pdf.
- O'DONNELL, T.J. 2009. « Chapter 7: Computer Representations of Molecular Structures ». In *Design and Use of Relational Databases in Chemistry*, 230. Boca Raton: CRC Press Taylor & Francis Group. https://books.google.fr/books?id=pTrwTKzECOC&pg=PA82&lpg=PA82&dq=smiles+forme+kekule*&source=bl&ots=Ug9ySMOAFj&sig=goyNPY3eBxK7vLd4fUTU5IS2Z6w&hl=fr&sa=X&ve

- d=0ahUKEwiU2v-E1ubbAhUR-6QKHakYc8MQ6AEIZTAK#v=onepage&q=smiles%20forme%20kekule*&f=false.
- PAL, Pallabi, Indrani MITRA, et Kunal ROY. 2014. « QSPR Modeling of Odor Threshold of Aliphatic Alcohols Using Extended Topochemical Atom (ETA) Indices ». *Croatica Chemica Acta* 87 (1): 29-37. <https://doi.org/10.5562/cca2284>.
- PARAVISINI, Laurianne, Chantal SEPTIER, Cédric MORETTON, Henri NIGAY, Gaëlle ARVISENET, Elisabeth GUICHARD, et Catherine DACREMONT. 2014. « Caramel odor: Contribution of volatile compounds according to their odor qualities to caramel typicality ». *Food Research International* 57: 79-88. <https://doi.org/10.1016/j.foodres.2014.01.009>.
- PARZEN, Emanuel. 1961. « On estimation of a probability density function and mode », n° 40.
- PATTE, François, et Pieter H. PUNTER. 1979. « Experimental assessment of human olfactory thresholds in air for some thiols and alkanes » 4 (4): 351-54.
- PATTERSON, Matthew Q., Joseph C. STEVENS, William S. CAIN, et J. Enrique COMETTO-MUNIZ. 1993. « Detection thresholds for an olfactory mixture and its three constituent compounds » 18 (6): 723-34. <https://doi.org/10.1093/chemse/18.6.723>.
- PICARD, Céline. 2013. « Peut-on imaginer remplacer l'analyse sensorielle des produits cosmétiques par des instruments de laboratoire ». présenté à HYGIA - Les rendez-vous de l'Académie de Pharmacie. http://www.acadpharm.org/dos_public/Pres_C._PICARD_Conference_HYGIA_AnP_05.06.2013.pdf.
- PIERRETTE, Marjorie, et Annie MOCH. 2009. « Étude des prédicteurs de la gêne olfactive aux abords d'un site industriel ». *Psychologie Française* 54 (3): 259-70. <https://doi.org/10.1016/j.psfr.2008.12.001>.
- PILLET, Davy, Maurice PILLET, et Alexandre ALBEROLA. 2018. « Tests de valeur aberrante ». Entreprise. ELLISTAT. 2018. <https://ellistat.com/guide-dutilisateur/statistiques-descriptives/tests-de-valeurs-aberrantes/>.
- PLAVSIC, Dejan, Sonja NIKOLIC, Nenad TRINAJSTIC, et Zlatko MIHALIC. 1993. « On the Harary index for the characterization of chemical graphs ». *Journal of Mathematical Chemistry* 12 (1): 235-250.
- POLISANO, Kévin. 2018. « Cours de statistiques (L1-MAP 201) ». Université de Grenoble. https://hal.archives-ouvertes.fr/cel-01787365/file/Cours_Statistiques_Polisano.pdf.
- POLSTER, Johannes, et Peter SCHIEBERLE. 2015. « Structure-Odor correlations in homologous series of alkanethiols and attempts to predict odor thresholds by 3D-QSAR studies ». *Journal of Agricultural and Food Chemistry* 63 (5): 1419-32. <https://doi.org/10.1021/jf506135c>.
- RAKOTOMALALA, Ricco. 2008. « Estimation de l'erreur de prédiction : les techniques de ré-échantillonnage ». Tutoriels Tanagra. Tutoriels Tanagra pour le Data Mining et la Data Science. 2008. https://eric.univ-lyon2.fr/~ricco/cours/slides/resampling_evaluation.pdf.
- RAMA, Paula. 2013. « Topological Indices: The Modified Schultz Index ». présenté à The 3rd Combinatorics Day, Lisbonne. <http://www.mat.uc.pt/~combdays/3rdAbstracts/PaulaRama3rdcombdays.pdf>.
- RAWAT, Suman, et O. P. SATI. 2013. « Quantitative structure property relationship study for prediction of Boiling point of aliphatic alkanes ». *International Journal of Pharmacy and pharmaceutical science* 5: 107-109.
- RAZALI, Nornadiah Mohd, et Yap Bee WAH. 2011. « Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests » 2 (1): 21-33.
- RINCON, Carlos Andrés, Amaury DE GUARDIA, Annabelle COUVERT, Dominique WOLBERT, Sophie LE ROUX, Isabelle SOUTREL, et Guillaume NUNES. 2019. « Odor concentration (OC) prediction based on odor activity values (OAVs) during composting of solid wastes and digestates ». *Atmospheric Environment* 201: 1-12. <https://doi.org/10.1016/j.atmosenv.2018.12.030>.

- ROBERTS, Stephen, et Richard EVERSON. 2001. *Independant Components : Principles and practice*. Cambridge: Cambridge University press.
- RODDIER, François. 2012. *Thermodynamique de l'évolution, un essai de thermo-bio-sociologie*. Parole. Le temps d'apprendre. Artignosc-surVerdon. <https://www.editions-parole.net/produit/thermodynamique-de-levolution-un-essai-de-thermo-bio-sociologie/>.
- RODRIGUEZ-NAVAS, C., R. FORTEZA, et V. CERDA. 2012. « Use of thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS) on identification of odorant emission focus by volatile organic compounds characterisation ». *Chemosphere* 89 (11): 1426-36. <https://doi.org/10.1016/j.chemosphere.2012.06.013>.
- ROGNON, Christian, et Lionel POURTIER. 2000. « Les odeurs dans l'environnement ». *Techniques de l'Ingénieur* G2900 V1. <https://www.techniques-ingenieur.fr/res/pdf/encyclopedia/tiag0-g2900-version1.pdf>.
- . 2014. *Les odeurs dans l'environnement.pdf*. Ed. Techniques Ingénieur.
- ROTHER, M., et B. THOMAS. 1963. « Aromastoffe des brotes ». *Zeitschrift für Lebensmittel-Untersuchung und Forschung* 119 (4): 302-310.
- ROUBY, Catherine, et Gilles SICARD. 1997. « Des catégories d'odeurs ? » In *Catégorisation et cognition : de la perception au discours*, Kimé, 59-81. Paris: Danièle DUBOIS. <https://www.cairn.info/categorisation-et-cognition-de-la-perception-au-di--978284174101X-page-59.htm>.
- ROUBY, Catherine, T. THOMAS-DANGUIN, Gilles SICARD, M. VIGOUROUX, T. JIANG, J. POITEVINEAU, et S. ISSANCHOU. 2005. « Influence du contexte sémantique sur la performance d'identification d'odeurs » 50 (2): 225-39. <https://doi.org/10.1016/j.psfr.2004.11.003>.
- ROY, Kunal, Supratik KAR, et Rudra Narayan DAS. 2015. « Statistical Methods in QSAR/QSPR ». In *A Primer on QSAR/QSPR Modeling*, 37-59. Cham: Springer International Publishing.
- ROYAL SOCIETY OF CHEMISTRY. 2015. « ChemSpider-Search and share chemistry ». ChemSpider-Search and share chemistry. 2015. <http://www.chemspider.com/Chemical-Structure.7012.html?rid=44c831ae-6694-4582-a90e-24e2fc4aa7ee>.
- RUNKLER, Thomas A. 2012. « Cluster Tendency Assessment ». In *Data Analytics : Models and Algorithms for Intelligent Data Analysis*, Springer Science & Business Media, 137. Allemagne: LEHRBUCH. <https://books.google.fr/books?id=PaylsdmgkklC&pg=PA118&lpg=PA118&dq=indice+de+hopkins+0.5&source=bl&ots=1IWqdLb5k7&sig=Qv7R8HZcm0unlB1gJPhCX2G-qSU&hl=fr&sa=X&ved=0ahUKEwivylTQkPnSAhVD2BoKHQGCAGlQ6AEIHTAA#v=onepage&q=indice%20de%20hopkins%200.5&f=false>.
- RUTH, Jon H. 1986. « Odor Thresholds and Irritation Levels of Several Chemical Substances : A Review ». *American Industrial Hygiene Association Journal* 47 (3): pp 142-151. <https://doi.org/10.1080/15298668691389595>.
- RUTLEDGE, Douglas N. 2016. « Independant component analysis : theory, application and difficulties ». présenté à AgroStat2016, Lausanne, Suisse.
- SAIKI, Katsunori. 2003. « Standard odor for selection of panel members ». <https://www.env.go.jp/en/air/odor/measure/>.
- SANDERSON, R.T. 1988. « Principles of electronegativity Part I. General nature ». *ACSPublication* 2 (65): 112. <https://doi.org/10.1021/ed065p112>.
- SAPORTA, Gilbert. 2011. *Probabilités, analyse des données et statistique*. 3ème. TECHNIP. <http://www.editionstechnip.com/en/catalogue-detail/149/probabilites-analyse-des-donnees-et-statistique.html>.
- SAULNIER, Natalia, et Christine ZANIN. 2003. « Le bruit comme facteur de nuisance à la qualité de la vie citadine ». *La ville, le bruit et le son* 78 (2): 121-28. <https://doi.org/10.4000/geocarrefour.197>.

- SAVENHED, R., H. BOREN, et A. GRIMVALL. 1985. « Stripping analysis and chromatographic sniffing for the source identification of odorous compounds in drinking water ». *Elsevier Science Publishers B.V* 328: 219-31. [https://doi.org/10.1016/S0021-9673\(01\)87392-1](https://doi.org/10.1016/S0021-9673(01)87392-1).
- SCHOLKOPF, Bernhard, et Alexander J. SMOLA. 2002. *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press. Londres, Angleterre. <https://mitpress.mit.edu/books/learning-kernels>.
- SCHUHFRIED, Erna, Eugenio APREA, Lucas CAPPELLIN, Christos SOUKOULIS, Roberto VIOLA, Tilmann D. MARK, Flavia GASPERI, et Franco BIASIOLI. 2012. « Desorption kinetics with PTR-MS: Isothermal differential desorption kinetics from a heterogeneous inlet surface at ambient pressure and a new concept for compound identification ». *International Journal of Mass Spectrometry* 314: 33-41. <https://doi.org/10.1016/j.ijms.2012.01.013>.
- SCHULTZ, Harry P. 1989. « Topological organic chemistry. 1. Graph theory and topological indices of alkanes ». *ACS Publications* 3 (29): 227-28. <https://doi.org/10.1021/ci00063a012>.
- SNEATH, Robert. 2003. « Quality Control of Olfactometry at SRI and in Europe ». <https://www.env.go.jp/en/air/odor/measure/>.
- SOUICI, Mohamed Lotfi, Leïla LOURICI, et Djelloul MESSADI. 2007. « Relation structure/rétention chromatographique de treize alkylnaphtalenes ». *Lebanese Science Journal* 8 (1): 63.
- STAUNE, Jean. 2015. *Les clés du futur*. PLON. Paris. <https://books.google.fr/books?id=l1ZxCQAAQBAJ&pg=PT55&lpg=PT55&dq=plus+une+structure+est+n%C3%A9gumentropique,+plus+elle+maximise+l'entropie+autour+d'elle&source=bl&ots=8PJCwRjqgg&sig=UaDyWOGN7NmwtUMSPxwWSZo7EU&hl=fr&sa=X&ved=0ahUKWir0K72hPnbAhX0haYKHV-6DiQQ6AEIKDAA#v=onepage&q=plus%20une%20structure%20est%20n%C3%A9gumentropique%20plus%20elle%20maximise%20l'entropie%20autour%20d'elle&f=false>.
- STRUBE, Andrea, Andrea BUETTNER, et Michael CZERNY. 2012. « Influence of chemical structure on absolute odour thresholds and odour characteristics of ortho- and para-halogenated phenols and cresols ». *John Wiley & Sons* 27 (4): 304-12. <https://doi.org/10.1002/ffj.3101>.
- STUIVER, Minze. 1958. « Biophysics of the Sense of Smell ». Groningen: University of Groningen. [https://www.rug.nl/research/portal/publications/biophysics-of-the-sense-of-smell\(24ce3138-3ff1-4cb7-94a9-d1485f46101c\).html](https://www.rug.nl/research/portal/publications/biophysics-of-the-sense-of-smell(24ce3138-3ff1-4cb7-94a9-d1485f46101c).html).
- SU, Guifu, Liming XIONG, et Ivan GUTMAN. 2013. « Harary Index of the K-Th Power of a Graph ». *Applicable Analysis and Discrete Mathematics* 7 (1): 94-105. <https://doi.org/10.2298/AADM121130024S>.
- SUCKER, K., R. BOTH, M. BISCHOFF, R. GUSKI, U. KRAMER, et G. WINNEKE. 2008. « Odor frequency and odor annoyance Part 2: dose-response associations and their modifications by hedonic tone ». *International Archives of Occupational and Environmental Health* 81 (6): 683-94. <https://doi.org/10.1007/s00420-007-0262-4>.
- SUFFET, I.H., et P. ROSENFELD. 2007. « The Anatomy of Odour Wheels for Odours of Drinking Water, Wastewater, Compost and the Urban Environment ». *Water Science and Technology* 55 (5): 335-44. <https://doi.org/10.2166/wst.2007.196>.
- TENENHAUS, Michel. 1998a. « Chapitre 6 - L'algorithme NIPALS ». In *La régression PLS - Théorie et pratique*, Technip, 139. Paris.
- . 1998b. « Chapitre 9.5 - Aide à l'interprétation ». In *La régression PLS - Théorie et pratique*, Technip, 139. Paris.
- . 1998c. *La régression PLS - Théorie et pratique*. Technip. Paris.
- TENENHAUS, Michel, Jean-Pierre GAUCHI, et C. MENARDO. 1995. « Régression PLS et applications » 43 (1): 7-63.
- TODESCHINI, Roberto, et Viviana CONSONNI. 2009a. « Balaban Distance Connectivity Index ». In *Molecular Descriptors for Chemoinformatics*, John Wiley & Sons, 1:39-43. *Methods and Principles in Medicinal Chemistry* 41. Germany: MANNHOLD Raimund;KUBINYI Hugo;

- FOLKERS Gerd.
<https://books.google.fr/books?id=6Zp7Yrqzv8AC&pg=PA428&lpg=PA428&dq=FLEXIBILITY+IN+DICES+molecular+descriptors+for+chemoinformatics&source=bl&ots=Jo2MFN2GGh&sig=m624tNJFWA9YGw5pZDNAjccmRzM&hl=fr&sa=X&ved=0ahUKEwjxr--XiLjSAhXE0xOKHX9KCAgQ6AEIJDAB#v=onepage&q=FLEXIBILITY%20INDICES%20molecular%20descriptors%20for%20chemoinformatics&f=false>.
- — —. 2009b. *Molecular Descriptors for Chemoinformatics*. John Wiley & Sons. Vol. I. 2 vol. Methods and Principles in Medicinal Chemistry 41. Germany: MANNHOLD Raimund; KUBINYI Hugo; FOLKERS Gerd.
<https://books.google.fr/books?id=6Zp7Yrqzv8AC&pg=PA428&lpg=PA428&dq=FLEXIBILITY+IN+DICES+molecular+descriptors+for+chemoinformatics&source=bl&ots=Jo2MFN2GGh&sig=m624tNJFWA9YGw5pZDNAjccmRzM&hl=fr&sa=X&ved=0ahUKEwjxr--XiLjSAhXE0xOKHX9KCAgQ6AEIJDAB#v=onepage&q=FLEXIBILITY%20INDICES%20molecular%20descriptors%20for%20chemoinformatics&f=false>.
- — —. 2009c. « Wiener Index ». In *Molecular Descriptors for Chemoinformatics*, John Wiley & Sons, 1:934-36. Methods and Principles in Medicinal Chemistry 41. Germany: MANNHOLD Raimund; KUBINYI Hugo; FOLKERS Gerd.
<https://books.google.fr/books?id=6Zp7Yrqzv8AC&pg=PA428&lpg=PA428&dq=FLEXIBILITY+IN+DICES+molecular+descriptors+for+chemoinformatics&source=bl&ots=Jo2MFN2GGh&sig=m624tNJFWA9YGw5pZDNAjccmRzM&hl=fr&sa=X&ved=0ahUKEwjxr--XiLjSAhXE0xOKHX9KCAgQ6AEIJDAB#v=onepage&q=FLEXIBILITY%20INDICES%20molecular%20descriptors%20for%20chemoinformatics&f=false>.
- TOROPOV, Andrey A., Alla P. TOROPOVA, Emilio BENFENATI, Giuseppina GINI, Danuta LESZCZYNSKA, et Jerzy LESZCZYNSKI. 2013. « CORAL: QSPR Model of Water Solubility Based on Local and Global SMILES Attributes ». *Chemosphere* 90 (2): 877-80.
<https://doi.org/10.1016/j.chemosphere.2012.07.035>.
- TOROPOV, Andrey A., Alla P. TOROPOVA, Luigi CAPPELLINI, Emilio BENFENATI, et Enrico DAVOLI. 2016. « Odor Threshold Prediction by Means of the Monte Carlo Method ». *Ecotoxicology and Environmental Safety* 133: 390-94. <https://doi.org/10.1016/j.ecoenv.2016.07.039>.
- TOROPOV, Andrey A., Alla P. TOROPOVA, Tomasz PUZYN, Emilio BENFENATI, Giuseppina GINI, Danuta LESZCZYNSKA, et Jerzy LESZCZYNSKI. 2013. « QSAR as a Random Event: Modeling of Nanoparticles Uptake in PaCa2 Cancer Cells ». *Chemosphere* 92 (1): 31-37.
<https://doi.org/10.1016/j.chemosphere.2013.03.012>.
- TOROPOVA, Alla P., Andrey A. TOROPOV, Emilio BENFENATI, Danuta LESZCZYNSKA, et Jerzy LESZCZYNSKI. 2015. « QSAR Model as a Random Event: A Case of Rat Toxicity ». *Bioorganic & Medicinal Chemistry* 23 (6): 1223-30. <https://doi.org/10.1016/j.bmc.2015.01.055>.
- TRANCHANT, Jean. 1996. « Chromatographie en phase gazeuse ». *Techniques de l'ingénieurs, Mesures - Analyses | Techniques d'analyse*, PE 1 485: 41.
- TUDURI, Ludovic. 2002. « Analyse de traces de composés organiques volatils dans l'air par microextraction sur phase solide ». Pau: Université de Pau et des pays de l'Adour.
- TUFFERY, Stéphane. 2008. « Data Mining & Statistique Décisionnelle.pdf ».
<http://data.mining.free.fr/cours/Descriptives.pdf>.
- — —. 2010. *Data mining et statistique décisionnelle: l'intelligence des données*. TECHNIP. Vol. 3ème édition.
https://books.google.fr/books?id=AyIYAA4a2kC&pg=PA8&lpg=PA8&dq=r%C3%A9gression+PLS+agroalimentaire+cosm%C3%A9tique&source=bl&ots=0Gx83qSf_M&sig=76mK2kLsG0lywPdcS7HWI5AdwOI&hl=fr&sa=X&ved=2ahUKEwioUqBlurcAhVsMewKHdIFBLEQ6AEwAHoECAAQAQ#v=onepage&q=r%C3%A9gression%20PLS%20agroalimentaire%20cosm%C3%A9tique&f=false.

- . 2011. « Agglomerative Hierarchical Clustering ». In *Data Mining and Statistics for Decision Making*, WILEY, 246-47. Chichester: John Wiley and Sons.
<https://books.google.fr/books?id=5MTBlxZUKiC&pg=PA246&lpg=PA246&dq=SPRSQ+cluster&source=bl&ots=ce4lnP7G9A&sig=o1hgT8uRAb0jaCBAfX4J5k-i3HU&hl=fr&sa=X&ved=0ahUKEwjuiOWm8ajRAhVS5GMKHdWRDHAQ6AEIPzAF#v=onepage&q=SPRSQ%20cluster&f=false>.
- UENO, Hiroyuki, Saeko AMANO, B. MERECKA, J. KOSMIDER, et others. 2009. « Difference in the odor concentrations measured by the triangle odor bag method and dynamic olfactometry ». *Water Science and Technology* 59 (7): 1339.
- UENO, Hiroyuki, Higuchi MASATO, Sukehisa TATSUICH, et Yosiharu IWASAKI. 2004. « Comparative Study of Two Methods for Olfactory Measurement: Triangular Odor Bag Method and Dynamic Olfactometry ». *Proceedings of the Water Environment Federation* 2004 (3): 30–37.
- US EPA. 1992. *Reference guide to odor thresholds for hazardous air pollutants listed in the clean air act amendements of 1990*. EPA.
- VAN GEMERT, L.J. 2011. *Odour Thresholds*. OPP-Oliemans Punter&Partners BV. Zeist, The Netherlands.
- VAN HARREVELD, AP. 2001. « From odorant formation to odour nuisance: new definitions for discussing a complex process, *Water Science and Technology* ». *Water Science and Technology* 44 (9): 9-15.
- VAPNIK, Vladimir N. 1995. *The Nature of Statistical Learning Theory*. Springer. New-York.
<https://www.springer.com/la/book/9780387987804>.
- VDI-3882-2. 1994. *Norme VDI 3882-2*.
- VENABLES, W.N., et B.D RIPLEY. 2002. *Modern Applied Statistics with S*. Vol. Fourth Edition. Statistics and computing. Springer. http://www.bagualu.net/wordpress/wp-content/uploads/2015/10/Modern_Applied_Statistics_With_S.pdf.
- VESIN, Aude, Guy BOUCHOUX, Etienne QUIVET, TEMIME-ROUSEL Brice, et Henri WORTHAM. 2012. « Use of the HS-PTR-MS for online measurements of pyrethroids during indoor insecticide treatments ». *Analytical and Bioanalytical Chemistry* 403 (7): 1907 - 21.
<https://doi.org/10.1007/s00216-012-6003-x>.
- VON RANSON, C., et H.D. BELITZ. 1992. « Untersuchungen zur Struktur-Aktivitätsbeziehung bei Geruchsstoffen. 2. Mitteilung: Wahrnehmungs- und Erkennungsschwellenwerte sowie Geruchsqualitäten gesättigter und ungesättigter aliphatischer Aldehyde » 195 (6): 515-22.
- WAINWRIGHT, Martin J. 2009. « Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using L1-Constrained Quadratic Programming (Lasso) ». *IEEE Transactions on Information Theory* 55 (5): 2183 - 2202. <https://doi.org/10.1109/TIT.2009.2016018>.
- WANG, Renxiao. 2007. *XLOGP3 v 3.2.0 User Manual* (version v 3.2.0). En. Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences.
- WARD, Joe H. Jr. 1963. « Hierarchical Grouping to Optimize an Objective Function ». *STOR* 58 (301): 236-44.
- WEININGER, David. 1988. « SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules ». *ACS Publications* 1 (28): 31 - 36.
- WICKHAM, Hadley, Winston CHANG, Lionel HENRY, Thomas Lin PEDERSEN, Kohske TAKAHASHI, Claus WILKE, Kara WOO, et RStudio. 2018. « Package 'Ggplot2' ». CRAN. <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>.
- WIENER, Harry. 1947. « Structural Determination of Paraffin Boiling Points ». *ACS Publications* 1 (69): 17 - 20. <https://doi.org/10.1021/ja01193a005>.
- WOLD, Svante, Michael SJOSTROM, et Lennart ERIKSSON. 2001. « PLS-Regression: A Basic Tool of Chemometrics ». *Elsevier* 58 (2): 109 - 30. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- WRIGHT, R.H. 1978. « The Perception of Odor Intensity: Physics or Psychophysics? » *Chemical Senses* 3 (1): 73-79.

- WU, C., J. LIU, P. ZHAO, M. PIRINGER, et G. SCHAUBERGER. 2016. « Conversion of the chemical concentration of odorous mixtures into odour concentration and odour intensity: A comparison of methods ». *Atmospheric Environment* 127: 283-92. <https://doi.org/>. <https://doi.org/10.1016/j.atmosenv.2015.12.051>.
- Wu, Chuandong, Jiemin Liu, Peng Zhao, Wenhui Li, Luchun Yan, Martin Piringer, et Günther Schauburger. 2017. « Evaluation of the chemical composition and correlation between the calculated and measured odour concentration of odorous gases from a landfill in Beijing, China ». *Atmospheric Environment* 164: 337-47. <https://doi.org/10.1016/j.atmosenv.2017.06.010>.
- WUNSCHE, Laurent, C. VUILLEUMIER, I. P. BYFIELD, Iain MAY, et Michael J. KEARNEY. 1995. « Scent Characterization : From Human Perception to Electronic Noses ». In , 3:295-313. Istanbul, Turkey: K.H.C. Baser (Ed.). https://www.researchgate.net/publication/288254434_Scent_Characterization_From_human_perception_to_electronic_noses.
- XIE, Yu-Long, Ji-Hong WANG, Yi-Zeng LIANG, Li-Xian SUN, Xin-Hua SONG, et Ru-Qin YU. 1993. « Robust Principal Component Analysis by Projection Pursuit ». *Journal of Chemometrics* 7 (6): 527-41. <https://doi.org/10.1002/cem.1180070606>.
- XU, Philippe, Franck DAVOINE, et Thierry DENOEU. 2014. « Transformation de scores SVM en fonctions de croyance ». In . Rouen, France: HAL. <https://hal.archives-ouvertes.fr/hal-00988584/document>.
- XU, Xuan, Feng LUAN, Huitao LIU, et Yuan GAO. 2012. « Prediction of the odor thresholds of oxygen and nitrogen containing heterocycles using the quantitative structure-property relationship approach ». *Asian Journal of Chemistry* 24 (9): 3842.
- YANG, Sung Bong. 2003. « A comparative study on odor regulation in Japan and Korea ». <https://www.env.go.jp/en/air/odor/measure/>.
- YANG, Yiting, et Linyuan LU. 2011. « The Randić index and the diameter of graphs ». *ELSEVIER*, n° 311: 1333-43. <http://www.sciencedirect.com/science/article/pii/S0012365X11001208>.
- YILAN, Luo, et ZENG RUTONG. 2015. « Package "Clustertend" ». CRAN. <https://cran.r-project.org/web/packages/fastICA/fastICA.pdf>.
- YUAN, Yunyun, Philip D. MOSIER, et Yan ZHANG. 2012. « Quantitative structure-property relationship (QSPR) model for predicting acidities of ketones ». *Journal of Biophysical Chemistry* 3 (1): 49-57. <https://doi.org/10.4236/jbpc.2012.31007>.
- ZARZO, Manuel. 2012. « Effect of Functional Group and Carbon Chain Length on the Odor Detection Threshold of Aliphatic Compounds ». *Sensors* 12 (12): 4105-12. <https://doi.org/10.3390/s120404105>.
- ZHOU, Bo. 2006. « Bounds for the Schultz molecular topological index ». *MATCH Communications in mathematical and in computer chemistry*, n° 56: 189-94.
- ZIEMER, Patrick D., James WOO, et Taki ANAGNOSTOU. 2000. « Study of odor qualification of solvents used in coating compositions », 1999 Roon Awards Competition Paper, 72 (907): 97-102.

Annexe 1 RESULTATS DU CALCUL DE LA FE ET DU TC DES ARTICLES

Articles	Nombre de SP aberrants selon le test Grubbs	Nombre de SP testés par le test de Grubbs	Total des SP mesuré par l'article	FE	TC
Eykman.1927	0	0	4	0	0
Fors.et.Fors.Olofsson.1984.1985	0	0	13	0	0
Berger.al.1985	0	0	11	0	0
Nijssen.1988	0	0	4	0	0
Laska.2010	0	0	6	0	0
Sävenhed.al.1985	0	1	8	0	12
Wünsche.al.1995	0	1	7	0	14
Huijter.1917	0	1	5	0	20
Kraft.Popaj.2004	0	1	5	0	20
Patte.Punter.et.Punter..1979	0	2	8	0	25
Baydar.al.1992a.1992b.1993	0	1	4	0	25
Stuiver.1958	0	10	25	0	40
Schley.1934	0	4	8	0	50
Cain.Gent.1991	0	2	4	0	50
Goldenberg.1967	0	3	5	0	60
Geier.1936	0	5	8	0	62
Hendriks.1979	0	5	8	0	62
Zimer.al.2000	0	11	17	0	65
Laska.Hudson.1991	0	9	13	0	69
Stone.1963c	0	5	7	0	71
CM.al.2004	0	5	7	0	71
Roos.al.1985	0	3	4	0	75
Cain.al.2010	0	3	4	0	75
CM.A.2009ab	0	7	9	0	78
CM.A.2010a	0	6	7	0	86
Stone.al.1972	0	10	11	0	91
Don.1986	0	13	14	0	93
Eriksson.al.1976	0	4	4	0	100
Ahlström.al.1986a	0	4	4	0	100
Hoshika.al.1993	0	6	6	0	100
Wise.al.2007.Miyazawa.al.2009a	0	4	4	0	100
CM.A.2008	0	4	4	0	100
CM.A.2010b	0	4	4	0	100
Hellmann.Small.1973.1974	1	36	93	3	39
Nagata.2003	3	103	220	3	47
Anon.1980	2	45	47	4	96
Jung.1936	1	17	19	6	89
Deadman.Prigg.1959	1	12	21	8	57
Lindell.1991	1	12	19	8	63
Naus.1982	2	21	22	10	95
Zwaardemaker.1914.1927	3	27	35	11	77
Moriguchi.al.1983	1	8	15	12	53
Hall.Anderson.1983	2	17	29	12	59
Dravnieks.al.1986	1	8	12	12	67
Passy.1892abc	3	24	35	12	69
Van.Anrooij.1931	1	8	9	12	89
Tempelaar.1913	3	18	31	17	58
Punter.1983	6	32	57	19	56
Amoore.1977	1	5	7	20	71
Dalton.al.2007	1	5	5	20	100
Bedborough.Trott.1979	2	9	11	22	82
Von.Ranson.Belitz.1992ab	3	13	99	23	13
Nagy.1991	11	47	77	23	61
Baldus.1936	3	12	16	25	75
Logtenberg.1978	1	4	5	25	80
Henning.1924	2	8	9	25	89
Köster.1971	4	13	18	31	72
Atanasova..al..2005	1	3	4	33	75
Williams.al.1977	1	3	4	33	75
Ohma.1922	2	6	6	33	100
Punter..1975..1979.	2	6	6	33	100
Laing.1975	2	5	6	40	83
CM.al.1999	7	17	24	41	71
Van.Thriel.al.2006	4	9	15	44	60
Dravnieks.1974	9	19	31	47	61
Molhave.al.2000	1	2	5	50	40
Stevens.Cain.1987a	1	2	4	50	50
May.1966	13	26	33	50	79
Moschandreas.Jones.1983	2	4	5	50	80
CM.C.et.CM..1994	15	29	42	52	69
Laska.Ringh.2010	4	6	6	67	100
Nishida.al.1979	6	8	8	75	100

Annexe 2 RESULTATS DE LA CLASSIFICATION DES ARTICLES

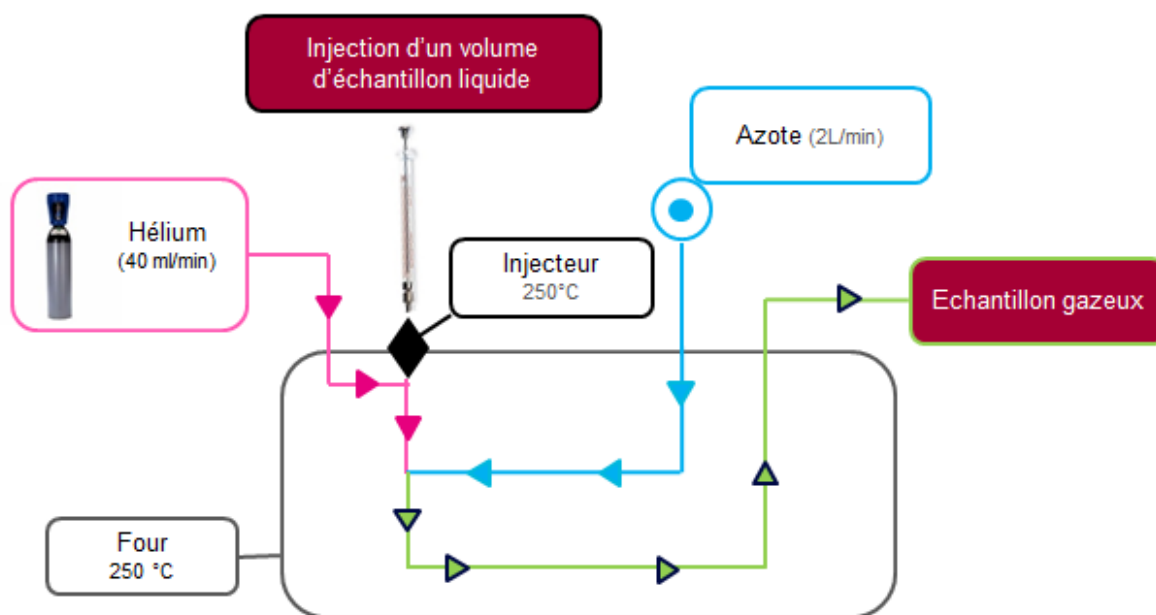
Articles	FE	TC	Classe
Ahlström.al.1986a	0	100	1
Cain.al.2010	0	75	1
Cain.Gent.1991	0	50	1
CM.A.2008	0	100	1
CM.A.2009ab	0	78	1
CM.A.2010a	0	86	1
CM.A.2010b	0	100	1
CM.al.2004	0	71	1
Don.1986	0	93	1
Eriksson.al.1976	0	100	1
Geier.1936	0	62	1
Goldenberg.1967	0	60	1
Hendriks.1979	0	62	1
Hoshika.al.1993	0	100	1
Laska.Hudson.1991	0	69	1
Roos.al.1985	0	75	1
Schley.1934	0	50	1
Stone.1963c	0	71	1
Stone.al.1972	0	91	1
Wise.al.2007.Miyazawa.al.2009a	0	100	1
Zimer.al.2000	0	65	1
Nagata.2003	3	47	1
Anon.1980	4	96	1
Jung.1936	6	89	1
Deadman.Prigg.1959	8	57	1
Lindell.1991	8	63	1
Naus.1982	10	95	1
Zwaardemaker.1914.1927	11	77	1
Dravnieks.al.1986	12	67	1
Hall.Anderson.1983	12	59	1
Moriguchi.al.1983	12	53	1
Passy.1892abc	12	69	1
Van.Anrooij.1931	12	89	1
Tempelaar.1913	17	58	2
Punter.1983	19	56	2
Amoore.1977	20	71	2
Dalton.al.2007	20	100	2
Bedborough.Trott.1979	22	82	2
Nagy.1991	23	61	2
Baldus.1936	25	75	2
Henning.1924	25	89	2
Logtenberg.1978	25	80	2
Köster.1971	31	72	2
Atanasova..al..2005	33	75	2
Ohma.1922	33	100	2
Punter..1975..1979.	33	100	2
Williams.al.1977	33	75	2
Laing.1975	40	83	3
CM.al.1999	41	71	3
Van.Thriel.al.2006	44	60	3
Dravnieks.1974	47	61	3
May.1966	50	79	3
Moschandreass.Jones.1983	50	80	3
Stevens.Cain.1987a	50	50	3
CM.C.et.CM..1994	52	69	3
Laska.Ringh.2010	67	100	3
Nishida.al.1979	75	100	3

Annexe 3 SEUILS DE PERCEPTION EXPERIMENTAUX : LE REFERENTIEL

N°CAS	Nom	Mes1	Mes2	Mes3	Mes4	Mes5	Mes6	Mes7	Mes8	Mes9	Mes10	Mes11	Mes12	Mes13	Mes14
108-88-3	toluène	1,3E+0	1,1E+0	1,2E+0	1,5E+0	1,4E+0	1,3E+0	1,2E+0	9,7E-1	1,2E+0	1,1E+0	1,1E+0	1,1E+0	NC	NC
71-36-3	n-butanol	1,8E-1	1,8E-1	2,1E-1	1,7E-1	1,0E-1	4,0E-2	5,0E-2	5,0E-2	4,0E-2	4,0E-2	1,2E-1	NC	NC	NC
123-86-4	acétate de butyle	1,1E-1	6,9E-2	5,0E-2	5,0E-2	7,0E-2	5,0E-2	5,0E-2	5,0E-2	5,5E-2	NC	NC	NC	NC	NC
120-92-3	cyclopentanone	8,7E-1	1,4E+0	9,6E-1	7,3E-1	8,3E-1	1,1E+0	1,2E+0	9,4E-1	1,4E+0	NC	NC	NC	NC	NC
470-82-6	eucalyptol	4,9E-3	4,9E-3	4,1E-3	1,9E-2	1,3E-3	9,6E-4	1,1E-3	1,2E-3	1,3E-3	NC	NC	NC	NC	NC
66-25-1	hexaNCI	2,4E-3	3,2E-3	6,3E-3	7,9E-3	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
623-42-7	méthyl butyrate	3,3E-2	2,8E-2	2,4E-2	3,2E-2	3,7E-2	1,6E-2	4,0E-2	NC	NC	NC	NC	NC	NC	NC
75-08-1	éthanethiol	1,4E-5	3,4E-5	2,5E-5	3,5E-5	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
5989-27-5	R-limonène	8,0E-2	8,0E-2	7,8E-2	5,5E-2	1,2E-1	8,8E-2	9,5E-2	1,0E-1	NC	NC	NC	NC	NC	NC
121-44-8	triéthylamine	4,9E-1	4,9E-1	4,5E-2	1,9E-2	2,8E-2	3,0E-2	3,1E-2	4,2E-2	4,1E-2	3,8E-2	4,3E-2	4,0E-2	NC	NC
431-03-8	2,3-butanedione	1,6E-4	9,9E-5	8,5E-5	1,1E-4	1,2E-4	1,2E-4	NC	NC	NC	NC	NC	NC	NC	NC
104-76-7	2-éthylhexanol	4,7E-1	5,5E-1	3,7E-1	3,5E-1	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
140-11-4	acétate de benzyle	1,9E-1	1,9E-1	1,4E-1	3,1E-2	2,9E-2	NC	NC	NC	NC	NC	NC	NC	NC	NC
80-56-8	alpha-pinène	1,8E-2	2,0E-2	2,0E-2	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
123-05-7	2-éthylhexaNCI	2,4E-1	2,1E-1	2,6E-1	1,6E-1	7,7E-2	6,8E-2	4,7E-2	NC	NC	NC	NC	NC	NC	NC
99-85-4	gamma-terpinène	4,7E-1	4,2E-1	4,7E-1	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
112-31-2	décaNCI	1,1E-2	9,5E-3	8,8E-3	1,1E-2	2,7E-3	3,3E-3	2,8E-3	1,8E-3	NC	NC	NC	NC	NC	NC
106-21-8	3,7-diméthyl-octanol	1,0E-2	1,0E-2	1,3E-2	4,1E-3	6,3E-3	3,7E-3	4,1E-3	NC	NC	NC	NC	NC	NC	NC
591-78-6	2-hexanone	6,7E-1	5,0E-1	6,2E-1	1,8E+0	5,7E-1	3,6E-1	NC	NC	NC	NC	NC	NC	NC	NC
589-38-8	3-hexanone	1,0E-1	9,0E-2	1,8E-1	1,2E-1	1,3E-1	8,0E-2	NC	NC	NC	NC	NC	NC	NC	NC
60-12-8	phényl-éthanol	2,0E-2	1,0E-2	1,5E-2	2,1E-2	2,5E-2	NC	NC	NC	NC	NC	NC	NC	NC	NC
110-54-3	hexane	5,5E+0	1,1E+1	2,4E+1	3,7E+1	1,3E+1	2,1E+1	NC	NC	NC	NC	NC	NC	NC	NC
592-41-6	1-hexène	4,1E-3	4,1E-3	6,2E-3	2,6E-3	1,9E-3	3,3E-4	3,3E-4	1,7E-4	1,8E-4	2,1E-4	NC	NC	NC	NC
1634-04-4	MTBE	4,0E-2	4,3E-2	1,1E-2	1,5E-2	2,9E-2	2,7E-2	2,5E-2	2,5E-2	1,5E-2	NC	NC	NC	NC	NC
64-17-5	éthanol	3,8E-1	3,7E-1	3,5E-1	2,6E-1	2,1E-1	1,8E-1	1,4E-1	8,1E-2	5,8E-2	6,1E-2	1,0E-1	1,2E-1	4,3E-2	6,1E-2
98-86-2	acétophenone	1,3E-2	1,6E-2	8,9E-3	2,0E-2	1,6E-2	1,3E-2	1,8E-2	1,1E-2	1,2E-2	1,9E-2	NC	NC	NC	NC
105-54-4	éthylbutyrate	4,3E-4	5,1E-4	5,8E-4	7,3E-4	6,0E-4	5,6E-4	5,8E-4	5,4E-4	NC	NC	NC	NC	NC	NC
624-24-8	Méthyl-valérate	9,0E-3	1,1E-2	9,0E-3	1,0E-2	1,2E-2	1,3E-2	1,4E-2	1,3E-2	NC	NC	NC	NC	NC	NC
590-86-3	3-méthylbutaNCI	6,1E-4	5,9E-4	3,9E-4	1,9E-4	2,4E-4	1,2E-4	NC	NC	NC	NC	NC	NC	NC	NC
75-18-3	DMS	4,7E-3	3,6E-3	3,7E-3	4,4E-3	9,9E-4	1,3E-3	1,9E-3	NC	NC	NC	NC	NC	NC	NC
67-64-1	acétone	2,4E+1	2,2E+1	3,0E+0	3,3E+0	6,0E+0	6,2E+0	7,6E+0	8,2E+0	7,9E+0	1,1E+1	8,1E+0	9,3E+0	NC	NC
78-93-3	2-butanone	1,6E+0	1,9E+0	2,4E+0	2,4E+0	2,8E+0	2,8E+0	NC	NC	NC	NC	NC	NC	NC	NC
141-78-6	acétate d'éthyle	2,9E+0	2,7E+0	2,6E+0	3,5E+0	4,9E+0	4,3E+0	NC	NC	NC	NC	NC	NC	NC	NC
111-66-0	1-octène	1,0E-2	1,0E-2	1,0E-2	1,0E-2	1,0E-2	1,0E-2	NC	NC	NC	NC	NC	NC	NC	NC
109-99-9	tétrahydrofurane	2,3E+0	2,2E+0	2,4E+0	3,4E+0	1,7E+0	2,0E+0	NC	NC	NC	NC	NC	NC	NC	NC
60-29-7	diéthyléther	1,1E-1	1,5E-1	1,0E-1	9,0E-2	1,4E-1	7,0E-2	NC	NC	NC	NC	NC	NC	NC	NC
96-17-3	2-méthylbutaNCI	1,8E-3	1,9E-3	5,5E-3	2,0E-3	1,2E-3	1,3E-3	3,4E-3	NC	NC	NC	NC	NC	NC	NC
592-76-7	1-heptène	9,0E-2	1,4E-1	1,5E-1	1,2E-1	2,0E-1	1,6E-1	NC	NC	NC	NC	NC	NC	NC	NC
105-37-3	éthyl propioNCte	2,0E-2	2,0E-2	2,0E-2	2,0E-2	2,0E-2	2,0E-2	4,0E-2	NC	NC	NC	NC	NC	NC	NC
108-10-1	4-méthyl-2-pentanone	5,2E-1	6,7E-1	7,2E-1	7,2E-1	7,2E-1	8,2E-1	NC	NC	NC	NC	NC	NC	NC	NC
100-42-5	styrène	6,0E-2	5,0E-2	9,0E-2	7,0E-2	6,0E-2	7,0E-2	NC	NC	NC	NC	NC	NC	NC	NC
110-83-8	cyclohexène	2,6E-1	2,5E-1	2,8E-1	2,2E-1	2,3E-1	1,5E-1	NC	NC	NC	NC	NC	NC	NC	NC
95-47-6	o-xylène	1,6E+0	1,6E+0	1,2E+0	9,8E-1	1,3E+0	1,5E+0	NC	NC	NC	NC	NC	NC	NC	NC
110-43-0	2-heptanone	9,0E-2	8,0E-2	9,0E-2	1,2E-1	6,0E-2	8,0E-2	NC	NC	NC	NC	NC	NC	NC	NC
108-83-8	2,6-diméthyl-4-heptanone	6,7E-1	5,4E-1	4,3E-1	3,8E-1	4,3E-1	3,4E-1	NC	NC	NC	NC	NC	NC	NC	NC

Annexe 4 PROTOCOLE DE VOLATILISATION D'UN ECHANTILLON LIQUIDE

Le principal intérêt de ce protocole est d'obtenir un échantillon gazeux de concentration en molécule(s) connue. Cela nécessite de constituer un échantillon liquide de composition connue en nature et quantité. Ensuite, le processus de volatilisation de l'échantillon liquide en échantillon gazeux est présenté ci-dessous.



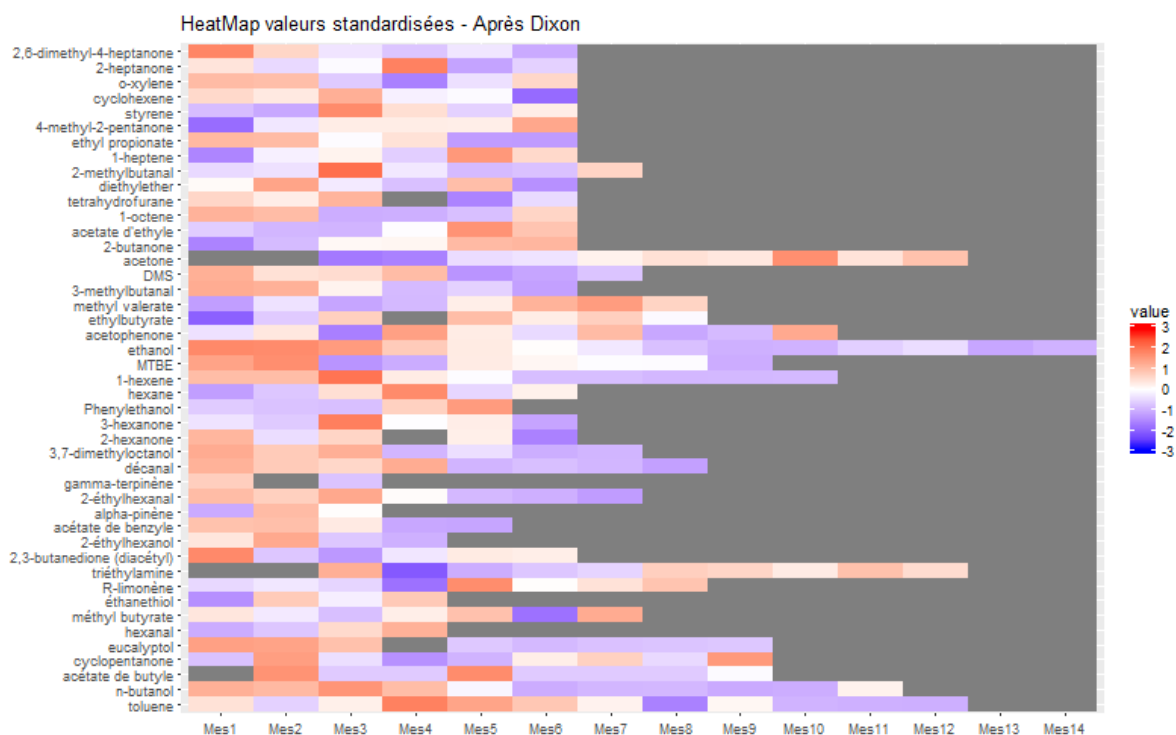
Le liquide est volatilisé dans l'injecteur du chromatographe chauffé à 250°C sous flux d'hélium. La phase gazeuse est ensuite diluée avec 40L d'azote à l'intérieur du four chauffé à 250°C pour obtenir le mélange gazeux synthétique dans un sac Nalophan®.

Tous les paramètres de la volatilisation sont contrôlés donc la concentration de l'échantillon gazeux est maîtrisée.

Deux cas de figures sont distingués : l'un vise à mesurer le seuil de perception d'une molécule (cas 1) alors que l'autre vise à fabriquer un mélange gazeux synthétique de plusieurs molécules (cas 2).

Dans le premier cas, l'échantillon liquide contient un corps pur et dans le second cas un mélange de molécules.

Annexe 5 HEATMAP DES VALEURS OBTENUES SUITE AU TEST DE DIXON SUR DONNEES DU REFERENTIEL NON STANDARDISEES



Annexe 6 Répétabilité des mesures expérimentales

N.CAS.1	Nom	Mesure 1	Mesure 2	Mesure 3	Mesure 4	Mesure 5	Mesure 6	Mesure 7	Mesure 8	Mesure 9	Mesure 10	Mesure 11	Mesure 12	Mesure 13	Mesure 14	Moyenne	Mode	Répétabilité
108-88-3	toluene	1,3E+00	1,1E+00	1,2E+00	1,5E+00	1,4E+00	1,3E+00	1,2E+00	9,7E-01	1,2E+00	1,1E+00	1,1E+00	1,1E+00	NC	NC	1,2E+00	1,2E+00	1,3E-01
71-36-3	n-butanol	1,8E-01	1,8E-01	2,1E-01	1,7E-01	1,0E-01	4,2E-02	5,0E-02	5,0E-02	4,0E-02	4,2E-02	1,2E-01	NC	NC	NC	1,1E-01	5,2E-02	6,4E-02
123-86-4	acétate de butyle	NC	6,9E-02	5,0E-02	5,0E-02	7,0E-02	5,0E-02	5,0E-02	5,0E-02	5,5E-02	NC	NC	NC	NC	NC	5,6E-02	5,0E-02	1,1E-02
120-92-3	cyclopentanone	8,7E-01	1,4E+00	9,6E-01	7,3E-01	8,3E-01	1,1E+00	1,2E+00	9,4E-01	1,4E+00	NC	NC	NC	NC	NC	1,1E+00	9,1E-01	2,7E-01
470-82-6	eucalyptol	4,9E-03	4,8E-03	4,1E-03	NC	1,3E-03	9,6E-04	1,1E-03	1,2E-03	1,3E-03	NC	NC	NC	NC	NC	2,5E-03	1,2E-03	2,2E-03
66-25-1	hexanal	2,4E-03	3,2E-03	6,3E-03	7,9E-03	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	4,9E-03	3,2E-03	7,6E-03
623-42-7	méthyl butyrate	3,3E-02	2,8E-02	2,4E-02	3,2E-02	3,7E-02	1,6E-02	4,0E-02	NC	NC	NC	NC	NC	NC	NC	3,0E-02	3,3E-02	1,1E-02
75-08-1	éthanthiol	1,4E-05	3,4E-05	2,5E-05	3,5E-05	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	2,7E-05	3,2E-05	2,8E-05
5989-27-5	R-limonène	7,8E-02	8,2E-02	7,8E-02	5,5E-02	1,2E-01	8,8E-02	9,5E-02	1,0E-01	NC	NC	NC	NC	NC	NC	8,7E-02	8,3E-02	2,3E-02
121-44-8	triéthylamine	NC	NC	4,5E-02	1,9E-02	2,8E-02	3,0E-02	3,1E-02	4,1E-02	4,1E-02	3,8E-02	4,3E-02	4,0E-02	NC	NC	3,6E-02	4,1E-02	8,5E-03
431-03-8	2,3-butanedione	1,6E-04	9,9E-05	8,5E-05	1,1E-04	1,2E-04	1,2E-04	NC	NC	NC	NC	NC	NC	NC	NC	1,2E-04	1,2E-04	4,2E-05
104-76-7	2-éthylhexanol	4,7E-01	5,5E-01	3,7E-01	3,5E-01	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	4,3E-01	3,8E-01	2,7E-01
140-11-4	acétate de benzyle	1,9E-01	1,9E-01	1,4E-01	3,1E-02	2,9E-02	NC	NC	NC	NC	NC	NC	NC	NC	NC	1,2E-01	1,7E-01	1,7E-01
80-56-8	alpha-pinène	1,8E-02	2,0E-02	1,9E-02	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	1,9E-02	1,9E-02	4,9E-03
123-05-7	2-éthylhexanal	2,4E-01	2,1E-01	2,6E-01	1,6E-01	7,7E-02	6,8E-02	4,7E-02	NC	NC	NC	NC	NC	NC	NC	1,5E-01	7,7E-02	1,2E-01
99-85-4	gamma-terpinène	4,7E-01	NC	4,7E-01	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	4,7E-01	4,7E-01	3,3E-02
112-31-2	décanal	1,1E-02	9,5E-03	8,8E-03	1,1E-02	2,7E-03	3,3E-03	2,8E-03	1,8E-03	NC	NC	NC	NC	NC	NC	6,4E-03	2,8E-03	5,1E-03
106-21-8	3,7-dimethyloctanol	1,3E-02	1,1E-02	1,3E-02	4,1E-03	6,3E-03	3,7E-03	4,1E-03	NC	NC	NC	NC	NC	NC	NC	8,0E-03	4,6E-03	6,2E-03
591-78-6	2-hexanone	6,7E-01	5,0E-01	6,2E-01	NC	5,7E-01	3,6E-01	NC	NC	NC	NC	NC	NC	NC	NC	5,5E-01	6,1E-01	2,5E-01
589-38-8	3-hexanone	1,0E-01	9,4E-02	1,8E-01	1,2E-01	1,3E-01	7,8E-02	NC	NC	NC	NC	NC	NC	NC	NC	1,2E-01	1,1E-01	5,8E-02
60-12-8	Phenylethanol	1,5E-02	1,5E-02	1,5E-02	2,1E-02	2,5E-02	NC	NC	NC	NC	NC	NC	NC	NC	NC	1,8E-02	1,5E-02	9,2E-03
110-54-3	hexane	5,5E+00	1,1E+01	2,4E+01	3,7E+01	1,3E+01	2,1E+01	NC	NC	NC	NC	NC	NC	NC	NC	1,9E+01	1,2E+01	1,9E+01
592-41-6	1-hexene	4,1E-03	4,1E-03	6,2E-03	2,6E-03	1,9E-03	3,3E-04	3,3E-04	1,7E-04	1,8E-04	2,1E-04	NC	NC	NC	NC	2,0E-03	5,2E-04	2,2E-03
1634-04-4	MTBE	4,0E-02	4,3E-02	1,1E-02	1,5E-02	2,9E-02	2,7E-02	2,5E-02	2,5E-02	1,5E-02	NC	NC	NC	NC	NC	2,5E-02	2,5E-02	1,2E-02
64-17-5	ethanol	3,8E-01	3,7E-01	3,5E-01	2,6E-01	2,1E-01	1,8E-01	1,4E-01	8,1E-02	5,8E-02	6,1E-02	1,0E-01	1,2E-01	4,3E-02	6,1E-02	1,7E-01	9,4E-02	9,8E-02
98-86-2	acetophenone	1,3E-02	1,6E-02	8,9E-03	2,0E-02	1,6E-02	1,3E-02	1,8E-02	1,1E-02	1,2E-02	1,9E-02	NC	NC	NC	NC	1,5E-02	1,3E-02	3,9E-03
105-54-4	ethylbutyrate	4,3E-04	5,1E-04	5,8E-04	7,3E-04	6,0E-04	5,6E-04	5,8E-04	5,4E-04	NC	NC	NC	NC	NC	NC	5,7E-04	5,8E-04	1,0E-04
624-24-8	methyl valerate	9,4E-03	1,1E-02	9,5E-03	9,9E-03	1,2E-02	1,3E-02	1,4E-02	1,3E-02	NC	NC	NC	NC	NC	NC	1,1E-02	1,0E-02	2,2E-03
590-86-3	3-methylbutanal	6,1E-04	5,9E-04	3,9E-04	1,9E-04	2,4E-04	1,2E-04	NC	NC	NC	NC	NC	NC	NC	NC	3,6E-04	2,2E-04	3,4E-04
75-18-3	DMS	4,7E-03	3,6E-03	3,7E-03	4,4E-03	9,9E-04	1,3E-03	1,9E-03	NC	NC	NC	NC	NC	NC	NC	2,9E-03	4,0E-03	2,1E-03
67-64-1	acetone	NC	NC	3,0E+00	3,3E+00	6,0E+00	6,2E+00	7,6E+00	8,2E+00	7,9E+00	1,1E+01	8,1E+00	9,3E+00	NC	NC	7,1E+00	7,9E+00	2,6E+00
78-93-3	2-butanone	1,6E+00	1,9E+00	2,4E+00	2,4E+00	2,8E+00	2,8E+00	NC	NC	NC	NC	NC	NC	NC	NC	2,3E+00	2,5E+00	8,1E-01
141-78-6	acetate d'ethyle	2,9E+00	2,7E+00	2,6E+00	3,4E+00	4,9E+00	4,3E+00	NC	NC	NC	NC	NC	NC	NC	NC	3,5E+00	2,9E+00	1,5E+00
111-66-0	1-octene	1,1E-02	1,1E-02	6,9E-03	7,0E-03	7,3E-03	1,0E-02	NC	NC	NC	NC	NC	NC	NC	NC	8,9E-03	7,2E-03	3,4E-03
109-99-9	tetrahydrofurane	2,3E+00	2,2E+00	2,4E+00	NC	1,7E+00	2,0E+00	NC	NC	NC	NC	NC	NC	NC	NC	2,1E+00	2,3E+00	5,5E-01
60-29-7	diethylether	1,1E-01	1,5E-01	1,0E-01	8,6E-02	1,4E-01	6,8E-02	NC	NC	NC	NC	NC	NC	NC	NC	1,1E-01	1,0E-01	5,3E-02
96-17-3	2-methylbutanal	1,8E-03	1,9E-03	5,5E-03	2,0E-03	1,2E-03	1,3E-03	3,4E-03	NC	NC	NC	NC	NC	NC	NC	2,5E-03	1,7E-03	2,1E-03
592-76-7	1-heptene	8,6E-02	1,4E-01	1,5E-01	1,2E-01	2,0E-01	1,6E-01	NC	NC	NC	NC	NC	NC	NC	NC	1,4E-01	1,4E-01	6,3E-02
105-37-3	ethyl propionate	2,3E-02	2,3E-02	2,0E-02	2,2E-02	1,7E-02	1,7E-02	NC	NC	NC	NC	NC	NC	NC	NC	2,0E-02	2,2E-02	5,2E-03
108-10-1	4-methyl-2-pentanone	5,2E-01	6,7E-01	7,2E-01	7,2E-01	7,2E-01	8,2E-01	NC	NC	NC	NC	NC	NC	NC	NC	7,0E-01	7,2E-01	1,6E-01
100-42-5	styrene	5,7E-02	5,3E-02	9,2E-02	7,5E-02	6,0E-02	7,2E-02	NC	NC	NC	NC	NC	NC	NC	NC	6,8E-02	5,9E-02	2,3E-02
110-83-8	cyclohexene	2,6E-01	2,5E-01	2,8E-01	2,2E-01	2,3E-01	1,5E-01	NC	NC	NC	NC	NC	NC	NC	NC	2,3E-01	2,4E-01	7,4E-02
95-47-6	o-xylene	1,6E+00	1,6E+00	1,2E+00	1,3E-01	1,3E+00	1,5E+00	NC	NC	NC	NC	NC	NC	NC	NC	1,3E+00	1,5E+00	3,8E-01
110-43-0	2-heptanone	9,5E-02	7,7E-02	8,5E-02	1,2E-01	6,3E-02	7,5E-02	NC	NC	NC	NC	NC	NC	NC	NC	8,7E-02	7,8E-02	3,5E-02
108-83-8	2,6-dimethyl-4-heptanone	6,7E-01	5,4E-01	4,3E-01	3,8E-01	4,3E-01	3,4E-01	NC	NC	NC	NC	NC	NC	NC	NC	4,6E-01	4,0E-01	2,0E-01

Annexe 7 Ecart absolu à la moyenne (EAM) individuels pour chaque molécule du référentiel après standardisation

N.CAS.1	Nom	Mesure 1	Mesure 2	Mesure 3	Mesure 4	Mesure 5	Mesure 6	Mesure 7	Mesure 8	Mesure 9	Mesure 10	Mesure 11	Mesure 12	Mesure 13	Mesure 14	moyenne	mode	n	EAM
108-88-3	toluene	4,2E-01	-5,5E-01	2,2E-01	1,8E+00	1,3E+00	8,2E-01	1,8E-01	-1,5E+00	1,2E-01	-9,0E-01	-9,4E-01	-9,5E-01	NC	NC	7,2E-16	1,4E-01	12	0,81
71-36-3	n-butanol	1,1E+00	1,0E+00	1,5E+00	9,6E-01	-1,2E-01	-9,8E-01	-8,6E-01	-8,6E-01	-1,0E+00	-9,8E-01	2,0E-01	NC	NC	NC	2,3E-17	-8,3E-01	11	0,88
123-86-4	acétate de butyle	NC	1,5E+00	-6,2E-01	-6,2E-01	1,6E+00	-6,2E-01	-6,2E-01	-6,2E-01	NC	NC	NC	NC	NC	NC	-4,8E-16	-5,7E-01	8	0,79
120-92-3	cyclopentanone	-7,3E-01	1,4E+00	-3,8E-01	-1,3E+00	-9,0E-01	2,4E-01	6,7E-01	-4,4E-01	1,5E+00	NC	NC	NC	NC	NC	-1,7E-16	-5,7E-01	9	0,84
470-82-6	eucalyptol	1,3E+00	1,3E+00	9,5E-01	NC	-7,0E-01	-7,6E-01	-7,3E-01	-7,1E-01	-7,0E-01	NC	NC	NC	NC	NC	-1,4E-17	-7,1E-01	8	0,90
66-25-1	hexanal	-1,2E+00	-4,4E-01	5,6E-01	1,1E+00	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	-6,9E-17	5,5E-01	4	0,81
623-42-7	méthyl butyrate	4,7E-01	-5,1E-01	-8,7E-01	3,6E-01	9,2E-01	-1,5E+00	1,2E+00	NC	NC	NC	NC	NC	NC	NC	1,5E-16	6,0E-01	7	0,84
75-08-1	éthanethiol	-9,1E-01	8,6E-01	-8,2E-01	8,7E-01	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	8,3E-17	7,7E-01	4	0,87
5989-27-5	R-limonène	-4,6E-01	-2,8E-01	-5,0E-01	-1,7E+00	1,6E+00	3,1E-02	4,2E-01	8,6E-01	NC	NC	NC	NC	NC	NC	-2,2E-16	-2,2E-01	8	0,73
121-44-8	triéthylamine	NC	NC	1,1E+00	-1,8E+00	-1,1E+00	-8,5E-01	-7,3E-01	7,4E-01	6,6E-01	4,0E-01	9,0E-01	5,8E-01	NC	NC	-2,1E-16	7,0E-01	10	0,88
431-03-8	2,3-butanedione	1,7E+00	-6,8E-01	-1,2E+00	-2,9E-01	3,0E-01	2,4E-01	NC	NC	NC	NC	NC	NC	NC	NC	-1,3E-16	3,2E-02	6	0,74
104-76-7	2-éthylhexanol	3,6E-01	1,2E+00	-6,7E-01	-9,4E-01	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	-3,5E-16	-5,8E-01	4	0,80
140-11-4	acétate de benzyle	8,9E-01	9,2E-01	3,1E-01	-1,1E+00	-1,1E+00	NC	NC	NC	NC	NC	NC	NC	NC	NC	-1,2E-16	6,8E-01	5	0,85
80-56-8	alpha-pinène	-1,0E+00	9,8E-01	3,6E-02	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	-1,3E-15	3,4E-01	3	0,68
123-05-7	2-éthylhexanal	9,7E-01	6,9E-01	1,3E+00	5,8E-02	-8,4E-01	-9,5E-01	-1,2E+00	NC	NC	NC	NC	NC	NC	NC	1,2E-17	-8,4E-01	7	0,85
99-85-4	gamma-terpinène	7,1E-01	NC	-7,1E-01	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	4,3E-14	7,1E-01	2	0,71
112-31-2	décanal	1,1E+00	7,7E-01	6,2E-01	1,2E+00	-1,0E+00	-5,4E-01	-1,0E+00	-1,1E+00	NC	NC	NC	NC	NC	NC	-4,2E-17	-8,9E-01	8	0,91
106-21-8	3,7-dimethyloctanol	1,2E+00	7,6E-01	1,1E+00	-8,8E-01	-3,8E-01	-9,7E-01	-8,8E-01	NC	NC	NC	NC	NC	NC	NC	-6,3E-17	-7,7E-01	7	0,89
591-78-6	2-hexanone	1,1E+00	-4,0E-01	6,1E-01	NC	2,4E-01	-1,5E+00	NC	NC	NC	NC	NC	NC	NC	NC	-1,7E-17	5,1E-01	5	0,77
589-38-8	3-hexanone	-3,3E-01	-6,3E-01	1,8E+00	-2,3E-02	2,6E-01	-1,1E+00	NC	NC	NC	NC	NC	NC	NC	NC	-2,4E-16	-2,1E-01	6	0,69
60-12-8	Phenylethanol	-6,1E-01	-7,3E-01	-7,7E-01	6,8E-01	1,4E+00	NC	NC	NC	NC	NC	NC	NC	NC	NC	0,0E+00	-6,5E-01	5	0,84
110-54-3	hexane	-8,6E-01	-2,1E-01	1,3E+00	-1,3E+00	4,7E-02	9,7E-01	NC	NC	NC	NC	NC	NC	NC	NC	8,8E-17	-4,0E-01	6	0,77
592-41-6	1-hexene	1,1E+00	1,1E+00	2,0E+00	-2,7E-01	-3,9E-01	-6,7E-01	-6,7E-01	-7,2E-01	-7,2E-01	-7,1E-01	NC	NC	NC	NC	1,1E-17	-6,0E-01	10	0,83
1634-04-4	MTBE	1,4E+00	1,7E+00	-1,1E+00	-9,0E-01	1,1E-01	-2,0E-02	-1,5E-01	-1,6E-01	-9,0E-01	NC	NC	NC	NC	NC	1,9E-17	-2,0E-01	9	0,73
64-17-5	ethanol	1,4E+00	1,3E+00	1,1E+00	1,3E+00	6,8E-01	3,2E-01	-1,1E-01	-8,0E-01	-1,1E+00	-1,0E+00	-5,4E-01	-3,1E-01	-1,2E+00	-1,0E+00	1,6E-17	-7,8E-01	14	0,87
98-86-2	acetophenone	-3,6E-01	3,5E-01	-1,5E+00	1,4E+00	2,8E-01	-4,4E-01	9,8E-01	-1,0E+00	-8,2E-01	1,2E+00	NC	NC	NC	NC	6,7E-17	-4,6E-01	10	0,84
105-54-4	ethylbutyrate	-1,6E+00	-7,0E-01	2,1E-01	1,9E+00	3,9E-01	-8,5E-02	2,1E-01	-3,2E-01	NC	NC	NC	NC	NC	NC	-9,4E-17	1,3E-01	8	0,68
624-24-8	methyl valérate	-1,2E+00	-3,4E-01	-1,1E+00	-8,3E-01	2,5E-01	1,1E+00	1,4E+00	6,3E-01	NC	NC	NC	NC	NC	NC	-4,9E-16	-7,5E-01	8	0,85
590-86-3	3-methylbutanal	1,2E+00	1,1E+00	3,4E-01	-8,7E-01	-7,7E-01	-9,8E-01	NC	NC	NC	NC	NC	NC	NC	NC	-4,6E-17	-7,8E-01	6	0,87
75-18-3	DMS	1,1E+00	5,3E-01	5,8E-01	9,5E-01	-1,1E+00	-1,1E+00	-9,5E-01	NC	NC	NC	NC	NC	NC	NC	-7,9E-17	7,6E-01	7	0,90
67-64-1	acetone	NC	NC	1,6E+00	-1,9E+00	-7,7E-01	-6,9E-01	-1,7E-01	7,3E-02	-1,8E-02	1,2E+00	4,6E-02	5,2E-01	NC	NC	1,6E-16	-4,8E-02	10	0,70
78-93-3	2-butanone	-1,5E+00	-8,1E-01	1,1E-01	1,4E-01	1,0E+00	1,1E+00	NC	NC	NC	NC	NC	NC	NC	NC	-4,0E-16	4,6E-01	6	0,77
141-78-6	acetate d'éthyle	1,1E+00	8,1E-01	8,0E-01	-1,1E+00	-7,1E-01	-8,8E-01	NC	NC	NC	NC	NC	NC	NC	NC	1,1E-16	8,5E-01	6	0,90
111-66-0	1-octene	1,1E+00	9,7E-01	-9,7E-01	-9,6E-01	-7,7E-01	6,1E-01	NC	NC	NC	NC	NC	NC	NC	NC	-1,9E-17	-8,5E-01	6	0,90
109-99-9	tetrahydrofurane	5,9E-01	2,6E-01	1,1E+00	NC	-1,5E+00	-4,4E-01	NC	NC	NC	NC	NC	NC	NC	NC	-2,2E-16	5,1E-01	5	0,77
60-29-7	diethylether	7,5E-02	1,3E+00	-2,5E-01	-7,5E-01	9,4E-01	-1,3E+00	NC	NC	NC	NC	NC	NC	NC	NC	1,8E-16	-3,3E-01	6	0,78
96-17-3	2-methylbutanal	-5,3E-01	-5,0E-01	1,9E+00	-4,8E-01	-6,5E-01	-6,2E-01	8,5E-01	NC	NC	NC	NC	NC	NC	NC	-6,3E-17	-5,6E-01	7	0,79
592-76-7	1-heptene	-1,5E+00	-1,7E-01	1,8E-01	-5,8E-01	1,5E+00	5,6E-01	NC	NC	NC	NC	NC	NC	NC	NC	1,2E-16	1,5E-02	6	0,74
105-37-3	ethyl propionate	1,0E+00	9,9E-01	-4,9E-02	4,2E-01	-1,2E+00	-1,2E+00	NC	NC	NC	NC	NC	NC	NC	NC	-9,3E-18	6,2E-01	6	0,81
108-10-1	4-méthyl-2-pentanone	-1,8E+00	-2,8E-01	2,7E-01	2,6E-01	2,5E-01	1,3E+00	NC	NC	NC	NC	NC	NC	NC	NC	0,0E+00	2,5E-01	6	0,68
100-42-5	styrene	-7,9E-01	-1,0E+00	1,6E+00	4,7E-01	-5,4E-01	2,6E-01	NC	NC	NC	NC	NC	NC	NC	NC	-1,8E-17	-6,3E-01	6	0,79
110-83-8	cyclohexene	5,6E-01	3,2E-01	1,2E+00	-1,8E-01	-6,2E-02	-1,8E+00	NC	NC	NC	NC	NC	NC	NC	NC	6,6E-17	9,4E-02	6	0,68
95-47-6	o-xylene	9,8E-01	9,4E-01	-6,3E-01	-1,5E+00	-3,6E-01	5,8E-01	NC	NC	NC	NC	NC	NC	NC	NC	-3,8E-16	6,7E-01	6	0,83
110-43-0	2-heptanone	3,9E-01	-4,5E-01	-6,4E-02	1,8E+00	-1,1E+00	-5,4E-01	NC	NC	NC	NC	NC	NC	NC	NC	9,5E-17	-4,0E-01	6	0,72
108-83-8	2,6-diméthyl-4-heptanone	1,7E+00	6,0E-01	-3,2E-01	-6,9E-01	-3,0E-01	-1,0E+00	NC	NC	NC	NC	NC	NC	NC	NC	-1,0E-16	-5,2E-01	6	0,77

Annexe 8 Valeurs DE Z_1, Z_2, Re, Z'_1, Z'_2 et Re'

N° CAS	Re	X1	X2	X3	Z1	Z2	Z1'=log10(1,5+Z1)	Z2'=log10(1,5+Z2)	Re'=log10(1,5+Re)
108-10-1	1.62E-01	7.80E-01	7.24E-01	7.21E-01	5.84E-02	3.02E-03	1.93E-01	1.77E-01	2.21E-01
591-78-6	2.46E-01	7.08E-01	7.08E-01	6.06E-01	1.02E-01	1.02E-01	2.05E-01	2.05E-01	2.42E-01
110-54-3	1.85E+01	5.30E+00	5.30E+00	1.24E+01	7.10E+00	7.10E+00	9.35E-01	9.35E-01	1.30E+00
624-24-8	2.18E-03	1.10E-02	1.10E-02	1.01E-02	9.32E-04	9.32E-04	1.76E-01	1.76E-01	1.77E-01
141-78-6	1.55E+00	3.31E+00	3.57E+00	2.91E+00	4.06E-01	6.64E-01	2.80E-01	3.35E-01	4.84E-01
60-12-8	9.25E-03	1.44E-02	4.90E-03	1.53E-02	8.88E-04	1.04E-02	1.76E-01	1.79E-01	1.79E-01
75-18-3	2.12E-03	5.61E-03	1.53E-03	4.00E-03	1.61E-03	2.47E-03	1.77E-01	1.77E-01	1.77E-01
95-47-6	3.84E-01	1.44E+00	8.90E-01	1.49E+00	5.34E-02	6.03E-01	1.91E-01	3.23E-01	2.75E-01
108-88-3	1.33E-01	2.43E+00	1.13E+00	1.22E+00	1.21E+00	8.40E-02	4.33E-01	2.00E-01	2.13E-01
123-86-4	1.10E-02	1.00E-02	5.25E-02	5.04E-02	4.04E-02	2.10E-03	1.88E-01	1.77E-01	1.79E-01
75-08-1	2.82E-05	7.35E-05	4.30E-05	3.24E-05	4.10E-05	1.05E-05	1.76E-01	1.76E-01	1.76E-01
590-86-3	3.41E-04	7.50E-04	3.50E-04	2.20E-04	5.30E-04	1.30E-04	1.76E-01	1.76E-01	1.76E-01
78-93-3	8.11E-01	5.59E+00	3.07E+00	2.54E+00	3.04E+00	5.21E-01	6.57E-01	3.06E-01	3.64E-01
111-66-0	3.43E-03	1.85E+00	4.60E-03	7.18E-03	1.85E+00	2.58E-03	5.25E-01	1.77E-01	1.77E-01
71-36-3	6.44E-02	4.21E-01	1.64E-01	5.21E-02	3.69E-01	1.12E-01	2.72E-01	2.07E-01	1.94E-01
470-82-6	2.23E-03	4.21E-02	3.13E-02	1.18E-03	4.09E-02	3.01E-02	1.88E-01	1.85E-01	1.77E-01
66-25-1	7.57E-03	2.94E-02	2.47E-02	3.25E-03	2.62E-02	2.15E-02	1.84E-01	1.82E-01	1.78E-01
121-44-8	8.55E-03	2.25E-01	1.54E-01	4.09E-02	1.84E-01	1.14E-01	2.26E-01	2.08E-01	1.79E-01
80-56-8	4.93E-03	3.05E-01	2.30E-01	1.92E-02	2.86E-01	2.11E-01	2.52E-01	2.33E-01	1.78E-01
99-85-4	3.29E-02	5.49E+01	2.50E+00	4.70E-01	5.44E+01	2.03E+00	1.75E+00	5.48E-01	1.86E-01
112-31-2	5.06E-03	5.67E-02	5.10E-02	2.81E-03	5.39E-02	4.81E-02	1.91E-01	1.90E-01	1.78E-01
64-17-5	9.78E-02	1.87E+01	1.11E+00	9.36E-02	1.86E+01	1.01E+00	1.30E+00	4.00E-01	2.04E-01
105-54-4	1.05E-04	1.70E-05	4.53E-05	5.78E-04	5.61E-04	5.33E-04	1.76E-01	1.76E-01	1.76E-01
67-64-1	2.58E+00	5.56E+01	2.96E+00	7.94E+00	4.77E+01	4.98E+00	1.69E+00	8.12E-01	6.10E-01
60-29-7	5.28E-02	7.70E-01	5.00E-01	1.00E-01	6.70E-01	4.00E-01	3.36E-01	2.79E-01	1.91E-01
110-43-0	3.53E-02	2.60E-01	1.28E-01	7.81E-02	1.82E-01	4.99E-02	2.26E-01	1.90E-01	1.86E-01
623-42-7	0.01139535	0.07691	0.077	0.03322	0.04368	0.04378	1.89E-01	1.89E-01	1.79E-01
431-03-8	4.2082E-05	0.00057	0.00193	0.00012	0.00046	0.00181	1.76E-01	1.77E-01	1.76E-01
592-41-6	2.23E-03	4.80E-01	4.80E-01	5.22E-04	4.79E-01	4.79E-01	2.97E-01	2.97E-01	1.77E-01
592-76-7	6.28E-02	1.50E+00	1.50E+00	1.43E-01	1.36E+00	1.36E+00	4.56E-01	4.56E-01	1.94E-01
105-37-3	0.00516979	1.19553	1.29751	0.02222	1.17331	1.27529	4.27E-01	4.43E-01	1.78E-01
100-42-5	0.02342956	0.11953	0.14198	0.0592	0.06033	0.08278	1.93E-01	1.99E-01	1.83E-01
110-83-8	7.44E-02	6.00E-01	6.00E-01	2.35E-01	3.65E-01	3.65E-01	2.71E-01	2.71E-01	1.97E-01

Annexe 9 Nombre d'observations pour chaque molécule de chaque jeu de données

N° CAS	Molécules	Nombre de valeurs du référentiel	Nombre de valeurs des données de la littérature	Nombre de valeurs des données de la littérature après nettoyage
108-88-3	toluene	12	17	10
71-36-3	Butan-1-ol	11	38	13
123-86-4	butyl acetate	8	14	6
120-92-3	cyclopentanone	9	1	0
470-82-6	2,2,4-trimethyl-3-oxabicyclo[2.2.2]octane (eucalyptol)	8	8	3
66-25-1	hexanal	4	10	6
623-42-7	methyl butanoate	7	3	1
75-08-1	éthanethiol	4	5	2
5989-27-5	6-methyl-3-prop-1-en-2-yl-7,8,9-trioxabicyclo[4.2.1]nonane (R-limonène)	8	3	0
121-44-8	N,N-diethylethanamine	10	3	2
431-03-8	butane-2,3-dione	6	4	3
104-76-7	2-ethylhexan-1-ol	4	2	0
140-11-4	benzyl acetate	5	1	0
80-56-8	4,6,6-trimethylbicyclo[3.1.1]hept-3-ene (alpha-pinène)	3	3	2
123-05-7	2-ethylhexanal	7	0	0
99-85-4	1-methyl-4-propan-2-ylcyclohexa-1,4-diene (gamma-terpinène)	2	2	1
112-31-2	decanal	8	3	2
106-21-8	3,7-dimethyloctan-1-ol	7	0	0
591-78-6	hexan-2-one	5	2	2
589-38-8	hexan-3-one	6	0	0
60-12-8	2-phenylethanol	5	6	1
110-54-3	hexane	6	1	1
592-41-6	hex-1-ene	10	1	1
1634-04-4	2-methoxy-2-methylpropane(MTBE)	9	2	0
64-17-5	ethanol	14	13	6
98-86-2	1-phenylethanone (acetophenone)	10	2	0
105-54-4	ethyl butanoate	8	7	3
624-24-8	methyl pentanoate	8	1	1
590-86-3	3-methylbutanal	6	2	1
75-18-3	methylsulfanylmethane(DMS)	7	7	2
67-64-1	propan-2-one (acetone)	10	17	7
78-93-3	butan-2-one	6	10	4
141-78-6	ethyl acetate	6	12	5
111-66-0	oct-1-ene	6	3	1
109-99-9	oxolane(THF)	5	2	0
60-29-7	ethoxyethane	6	4	1
96-17-3	2-methylbutanal	7	0	0
592-76-7	hept-1-ene	6	1	1
105-37-3	ethyl propanoate	6	3	2
108-10-1	4-methylpentan-2-one	6	10	5
100-42-5	styrene	6	10	5
110-83-8	cyclohexene	6	1	1
95-47-6	1,2-xylene	6	6	3
110-43-0	heptan-2-one	6	11	6
108-83-8	2,6-dimethylheptan-4-one	6	2	0

Annexe 10 CALCUL DES INDICES SELECTIONNES POUR LE MODELE RSQA/RSQP

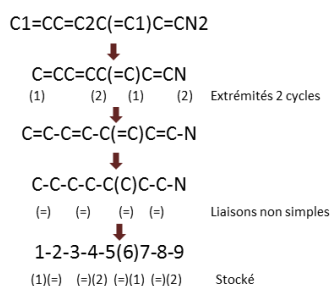
Les principales étapes du décodage du SMILES pour calculer les indices (ou variables d'entrée) sont présentées en Figure 69 pour l'exemple de l'indole. L'identifiants InChI est utilisé en parallèle pour notamment comptabiliser les atomes d'hydrogène et extraire l'ordre d'apparition des atomes de la formule brute.

Exemple : l'Indole (CAS=120-72-9)

InChI = 1S/C8H7N/c1-2-4-8-7(3-1)5-6-9-8/h1-6,9H

SMILES = C1=CC=C2C(=C1)C=CN2

Simplification de la chaîne SMILES (Principales étapes)



Numérotation de la molécule

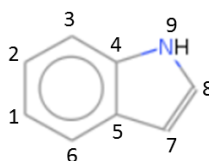


Figure 69 : Étapes du décodage de la grandeur SMILES de l'indole

Le processus présenté en Figure 69 consiste à utiliser la grandeur SMILES afin de localiser les éléments clés et de transformer cette grandeur en un codage en format standard InChI. En effet, si seule l'InChI est utilisé, il est par exemple impossible de déceler la présence et la localisation des doubles liaisons et des cycles. Le calcul des indices de composition et les indices constitutionnels présentés plus haut est possible à ce stade.

Concernant les indices topologiques et l'indice électrotopologique ZEP, la construction de matrices intermédiaires est nécessaire. Ces matrices sont considérées comme des graphes où les atomes sont des sommets et les liaisons des arrêtes [(Zoița BERINDE 2006) ; (BRUCKLER 2008) ; (Zoița BERINDE et BERINDE 2009a)]. Ces matrices ont connu une évolution au cours du temps. Au début, le calcul de ces indices était uniquement possible pour des molécules ayant uniquement des liaisons simples et sans hétéroatomes. Il s'agit de la matrice de proximité (A), de la matrice de distance (D) et de la réciproque de D (RD) [(PLAVSIC et al. 1993) ; (IVANCIUC, BALABAN, et BALABAN 1993) ; (DIUDEA et al. 1997)] dont le calcul est défini en Figure 70.

$$A_{ij} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ adjacents} \\ 0 & \text{sinon} \end{cases}$$

Matrice de proximité

$$D_{ij} = \begin{cases} d_{ij} & \text{si } i \neq j \\ 0 & \text{sinon} \end{cases}$$

Matrice de distance

$$RD_{ij} = \begin{cases} d_{ij}^{-1} & \text{si } i \neq j \\ 0 & \text{sinon} \end{cases}$$

Réciproque de D

Figure 70 : Matrices utilisées pour le calcul des indices topologiques - Cas des molécules avec liaisons simples et sans hétéroatomes

Le calcul de ces matrices a ensuite été adapté au cas des molécules contenant des hétéroatomes (NIKOLIC, TRINAJSTIC, et MIHALIC 1993). Cette nécessité d'extension est soulignée par Berinde en constatant que si la matrice A ne connaît pas d'extension alors les molécules isobutane, isobutène, 2-propanol, 2-chloropropane, 2-aminopropane seront associés à un même graph (Zoița BERINDE 2006). Cela est également nécessaire pour la matrice D et donc la matrice RD. La constitution des matrices étendues est présentée en Figure 71 [(NIKOLIC, TRINAJSTIC, et MIHALIC 1993) ; (MOZRZYMAS 2011)].

	Pondération des sommets	Pondération des branches
Paramètres définis pour l'extension	$(V_z^*)_i = 1 - \frac{6}{Z_i}$	$(E_z^*)_k = \frac{36}{b_k \cdot Z_i Z_j}$

$$Ah_{ij} = \begin{cases} V^* & \text{si } i \text{ hétéroatome} \\ E^* & \text{si } i \text{ et } j \text{ adjacents} \\ 0 & \text{sinon} \end{cases}$$

Matrice de proximité étendue

$$Dh_{ij} = \begin{cases} \sum_k (EZ^*)_k & \text{si } i \neq j \\ (V_z^*)_i & \text{si } i = j \end{cases}$$

Matrice de distance étendue

$$RDh_{ij} = \begin{cases} (\sum_k (EZ^*)_k)^{-1} & \text{si } i \neq j \\ (V_z^*)_i & \text{si } i = j \end{cases}$$

Réciproque de D étendue

Figure 71 : Matrices étendues utilisées pour construction des indices topologiques - Cas des molécules avec liaisons tous types de liaisons et avec hétéroatomes

Tous les indices, hormis XlogP3, sont calculés à partir de ces matrices. Le calcul des indices topologiques et électro-topologiques sont définis dans la littérature sur la base des matrices non étendues aux hétéroatomes :

- **Wiener** en 1947, noté W, traduit la distance globale entre atomes et correspond à la somme des termes de la matrice D [(WIENER 1947) ; (PLAVSIC et al. 1993) ; (GUTMAN et KLAUZAR 1997) ; (DOBRYNIN 1999) ; (TODESCHINI et CONSONNI 2009c) ; (HUA 2009)] ;
- **Zagreb** en 1972, noté M1 et M2, implique la valence des molécules [(GUTMAN et TRINAJSTIC 1972) ; (NESTEROV et al. 2004) ; (RAMA 2013)] ;
- **Randic et conductivité moléculaires** en 1975 [(NESTEROV et al. 2004) ; (BRUCKLER 2008) ; (Y. YANG et LU 2011)] ;

- **Balaban** en 1982, noté J, est un indice de connectivité de la molécule. Il est issu de la matrice D (ou Dh) et inclut la prise en compte du nombre de cycle et du nombre de liaisons total. Une version de cet indice, noté Jt, inclut le nombre de liaisons de chaque atome avec les autres [(BALABAN 1985) ; (TODESCHINI et CONSONNI 2009a)] ;
- **Schultz** en 1989, noté MTI, correspond à la somme de A et de D multiplié par la somme des lignes de A [(SCHULTZ 1989) ; (GUTMAN 1994) ; (GUTMAN et KLAVZAR 1997) ; (ZHOU 2006) ; (HUA 2009)] ;
- **Harary** en 1991, noté H, correspond à la somme des termes de la matrice RD [(PLAVSIC et al. 1993) ; (SU, XIONG, et GUTMAN 2013)] ;
- **Gutman** en 1994 est une version de l'indice de Schultz modifié puisque A et D ne sont pas additionnées mais multipliées [(GUTMAN 1994) ; (RAMA 2013)].

Dans le script de calcul mis en place, ils ont été appliqués sur les matrices étendues aux hétéroatomes.

Puis, Berinde (Zoita BERINDE et BERINDE 2004) a mis en place le calcul d'une autre matrice dérivée de la matrice Ah pour souligner le poids de la distance électronique qu'il a testé sur différentes familles de molécules [(Zoița BERINDE 2006) ; (Zoita BERINDE 2008) ; (Zoita BERINDE et BERINDE 2009b)]. Il s'agit de la matrice CEP pour « weighted electronic connectivity matrix » dont le calcul est présenté en Figure 72.

$$\boxed{\text{CEP}} = \begin{cases} w.e.d(i,j) & \text{si } i \text{ et } j \text{ adjacents} \\ 0 & \text{sinon} \end{cases} \quad \text{avec} \quad w.e.d(i,j) = \frac{1}{b_r} \cdot \frac{Z'_i + Z'_j}{v_i \cdot v_j}$$

et $Z'_i = Z_i \cdot v_i$

Figure 72 : Définition de la matrice CEP (« weighted electronic connectivity matrix »)

Cette matrice permet ensuite de calculer l'indice ZEP (Zoita BERINDE et BERINDE 2004) défini en Équation 11.

$$ZEP = \sum_{i=1}^N \left(\sum_{j=1}^N [CEP]_{ij} \right)^{1/2}$$

Avec N le nombre de lignes/colonnes de la matrice CEP
 i les lignes de la matrices CEP
 j les colonnes de la matrice CEP

Équation 11 : Calcul de l'indice topologique ZEP

L'indice électrotopologique d'état (E-State), noté S_i , est un indice mis en place pour tenir compte de l'environnement topologique et des interactions électroniques entre atomes d'une molécule [(KIER et HALL 1990) ; (HALL, MOHNEY, et KIER 1991) ; (HALL et KIER 1995)]. Il est défini par atome selon l'Équation 12 (EduSoft LC, s. d.).

$$S_i = I_i + \Delta I_i \quad \text{Avec} \quad \begin{array}{l} I_i : \text{la fonction d'état intrinsèque de l'atome } i \\ \Delta I_i : \text{la perturbation créée sur l'atome } i \text{ par son environnement} \end{array}$$

Équation 12 : Définition de l'indice électrotopologique d'état S_i par atome i

Il faut donc calculer la valeur d'état intrinsèque I_i et la perturbation ΔI_i de chaque atome i par molécule comme définis respectivement en Équation 13 et Équation 14.

$$I_i = \frac{(2/N_i)^2 \times \delta^v(i) + 1}{\delta(i)} \quad \text{Avec} \quad \delta(i) = \delta(i) = \sigma_i - H_i \quad \text{Où} \quad \begin{array}{l} \sigma_i : \text{nombre d'électron sur l'orbitale sigma de l'atome } i \\ H_i : \text{le nombre de H liés de l'atome } i \\ H_i : \text{le nombre de H liés de l'atome } i \\ Z_i^v : \text{le nombre d'électrons de valence de l'atome } i \\ Z_i : \text{le numéro atomique de l'atome } i \\ N_i : \text{le nombre quantique principal de l'atome } i \end{array}$$

$$\delta^v(i) = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1} \quad \text{Où}$$

N : le nombre quantique principal

Équation 13 : Définition de la valeur d'état intrinsèque d'un atome i

$$\Delta I_i = \sum_{j=1}^N \frac{I_i - I_j}{I_{ij}^2} \quad \text{Avec} \quad \begin{array}{l} I_i : \text{la valeur d'état intrinsèque de l'atome } i \\ I_j : \text{la valeur d'état intrinsèque de chaque atome } j \\ I_{ij} : \text{le nombre d'atome du plus court chemin entre l'atome } i \text{ et l'atome } j \end{array}$$

Équation 14 : Définition de la perturbation d'un atome i

Afin d'obtenir une valeur global de la perturbation des atomes, il est calculé dans la fonction R, l'indice E-State global noté « SiTot » et correspondant à la moyenne des S_i de la molécule.

Enfin, l'électronégativité de Sanderson est calculée en moyennant l'électronégativité respective de chaque atome [(SANDERSON 1988) ; (KnowledgeDoor LLC 2017)].

L'un des inconvénients de cette fonction est que la structure 3D n'est pas prise en compte, mais une description large des molécules est malgré tout utilisable. De plus, pour des raisons de structure et de propriétés intrinsèques, les indices ne peuvent pas être calculés pour les molécules chargées et les molécules silylées. Ce qui correspond à 15 molécules de la totalité inventoriée.

Une extension des indices de Wiener et de Balaban est calculée. Il s'agit du calcul de ces indices sur la matrice CEP. Ces trois indices sont ajoutés à la base.

Esters	106-63-8	2.2	7	12	0	0	0	0	0	20	6	2	0	0	5	1	1	0	128	6	2	0	85	52	50	3	19	4	18	3	4	9	313	23	550	7	38	92	41.5	8	19	3	26	27
Esters	140-88-5	1.3	5	8	0	0	0	0	0	14	4	2	0	0	4	1	0	0	100	7	2	0	38	42	40	3	14	3	12	2	4	10	141	17	206	5	23	57	29.5	5	13	3	19	21
Esters	80-62-6	1.4	5	8	0	0	0	0	0	14	4	2	0	0	3	0	1	0	100	7	2	0	36	46	46	3	14	3	13	2	4	11	133	18	213	5	23	57	31.25	4	12	3	20	21
Esters	96-33-3	0.8	4	6	0	0	0	0	0	11	3	2	0	0	2	1	0	0	86	7	2	0	22	38	36	2	11	3	10	2	4	10	85	14	115	4	17	43	24.5	4	10	3	16	18
Ethers	1708-29-8	0.5	4	6	0	0	0	0	0	11	4	1	0	0	2	2	0	0	70	6	2	1	12	30	29	2	10	3	9	2	3	6	67	10	90	5	7	16	26	1	3	2	16	15
Soufrés	10152-76-8	1.5	4	8	0	0	0	0	0	12	3	1	0	0	2	1	0	0	88	7	1	0	12	22	18	2	12	2	12	3	4	7	48	13	51	3	14	27	37.5	2	4	2	18	14
Soufrés	2179-57-9	2.2	6	10	0	0	0	0	0	17	5	2	0	0	4	2	0	0	146	8	2	0	45	42	36	3	16	4	16	5	5	11	169	33	223	6	36	80	53	5	13	2	27	22
Soufrés	592-88-1	2.2	6	10	0	0	0	0	0	16	4	2	0	0	4	2	0	0	114	7	2	0	35	38	32	3	16	3	15	3	4	9	133	20	180	5	24	56	37	5	11	2	21	19
Terpènes	127-91-3	3.1	10	16	0	0	0	0	0	27	10	1	0	0	3	3	3	1	136	5	3	2	98	68	82	4	23	5	23	4	2	6	461	26	1083	11	17	44	53.25	4	10	2	32	27
Terpènes	138-86-3	3.4	10	16	0	0	0	0	0	26	8	2	0	0	3	4	3	0	136	5	3	1	109	68	70	4	24	5	23	4	2	6	458	26	960	10	23	59	48	5	13	2	30	27
Terpènes	5989-08-2	4.6	15	24	0	0	0	0	0	41	16	1	0	0	4	5	4	2	204	5	4	3	292	104	130	7	35	7	35	7	2	6	1329	49	3369	17	30	76	84.75	6	16	2	50	41
Terpènes	5989-27-5	3.4	10	16	0	0	0	0	0	26	8	2	0	0	3	4	3	0	136	5	3	1	109	68	70	4	24	5	23	4	2	6	458	26	960	10	23	59	48	5	13	2	30	27
Terpènes	7785-26-4	2.8	10	16	0	0	0	0	0	27	10	1	0	0	3	3	3	1	136	5	3	2	99	70	84	4	23	5	23	4	2	6	461	26	1086	11	17	44	54.75	3	9	2	33	27
Terpènes	7785-70-8	2.8	10	16	0	0	0	0	0	27	10	1	0	0	3	3	3	1	136	5	3	2	99	70	84	4	23	5	23	4	2	6	461	26	1086	11	17	44	54.75	3	9	2	33	27
Terpènes	78-79-5	2.5	5	8	0	0	0	0	0	12	2	2	0	0	3	1	1	0	68	5	2	0	14	34	30	2	12	2	11	2	3	9	57	9	87	4	9	24	15.75	3	7	2	12	14
Terpènes	80-56-8	2.8	10	16	0	0	0	0	0	27	10	1	0	0	3	3	3	1	136	5	3	2	99	70	84	4	23	5	23	4	2	6	461	26	1086	11	17	44	54.75	3	9	2	33	27
Terpènes	99-85-4	2.8	10	16	0	0	0	0	0	26	8	2	0	0	3	4	3	0	136	5	3	1	108	70	74	4	24	5	23	4	3	7	450	26	942	10	23	59	49.5	5	13	2	31	27
Alcools	10415-87-9	3	12	18	0	0	0	0	0	31	7	0	0	6	2	8	2	0	178	6	4	1	239	92	98	5	28	6	26	5	2	7	973	40	2085	13	40	105	59.34	10	29	2	38	37
Aromatiques	106-44-5	1.9	7	8	0	0	0	0	0	16	2	0	0	6	1	5	1	0	108	7	4	1	44	70	74	3	14	4	13	3	3	9	202	23	374	8	15	45	30.16	4	13	2	22	23
Aromatiques	108-39-4	2	7	8	0	0	0	0	0	16	2	0	0	6	1	5	1	0	108	7	4	1	44	70	74	3	14	4	13	3	3	10	199	23	369	8	15	45	30.16	4	13	2	22	23
Aromatiques	108-95-2	1.5	6	6	0	0	0	0	0	13	1	0	0	6	0	6	0	0	94	7	4	1	29	62	64	2	11	3	10	2	3	9	140	19	233	7	12	36	23.34	4	12	3	18	20
Aromatiques	60-12-8	1.4	8	10	0	0	0	0	0	19	3	0	0	6	1	6	1	0	122	6	4	1	74	70	74	3	17	4	15	3	3	8	321	25	577	9	20	55	35.84	6	17	2	25	26
Aromatiques	89-83-8	3.3	10	14	0	0	0	0	0	25	5	0	0	6	3	4	3	0	150	6	4	1	119	88	99	4	22	5	21	4	3	9	488	35	1032	11	28	78	48.99	7	21	2	32	31
Aromatiques	95-48-7	2	7	8	0	0	0	0	0	16	2	0	0	6	1	5	1	0	108	7	4	1	43	70	75	3	14	4	13	3	3	10	197	23	361	8	15	45	30.17	4	13	2	22	23
Esters	138-22-7	1.1	7	14	0	0	0	0	0	23	8	1	0	0	4	3	0	0	146	6	1	0	123	48	48	4	22	5	20	4	4	8	450	26	781	8	48	105	55.67	7	17	3	33	30
Esters	25265-77-4	3.1	12	24	0	0	0	0	0	38	13	1	0	0	8	1	2	1	216	6	1	0	356	78	84	6	36	7	35	6	5	11	1298	47	2848	13	102	233	85.67	16	38	2	49	44
Alcools	110-80-5	-0.3	4	10	0	0	0	0	0	15	5	0	0	0	4	0	0	0	90	6	0	0	30	18	16	3	15	3	14	2	3	5	109	11	150	5	16	28	39	2	3	2	22	18
Alcools	111-76-2	0.8	6	14	0	0	0	0	0	21	7	0	0	0	4	2	0	0	118	5	0	0	75	26	24	4	20	4	19	3	3	5	273	16	425	7	29	53	51	4	7	2	28	24
Ethers	5131-66-8	1.1	7	16	0	0	0	0	0	24	8	0	0	0	4	3	0	0	132	5	0	0	102	32	30	4	23	4	22	4	3	6	373	19	647	8	36	68	56.67	5	9	2	32	27
Azotés	104-90-5	2	8	11	0	0	0	0	0	20	3	0	0	6	3	4	1	0	121	6	4	1	68	74	80	3	18	4	17	3	3	9	288	26	515	8	22	61	36.05	5	16	2	25	25
Azotés	108-48-5	1.7	7	9	0	0	0	0	0	17	2	0	0	6	2	5	0	0	107	6	4	1	44	70	74	3	15	4	14	3	3	10	194	23	340	7	17	49	30	4	13	2	22	22
Azotés	109-06-8	1.1	6	7	0	0	0	0	0	14	1	0	0	6	2	4	0	0	93	7	4	1	29	62	64	2	12	3	11	2	3	9	138	19	217	6	14	40	23.23	4	12	2	18	20
Azotés	110-86-1	0.7	5	5	0	0	0	0	0	11	0	0	0	6	2	3	0	0	79	7	4	1	17	54	54	2	9	3	8	2	3	10	89	16	119	5	11	32	16.46	3	10	2	14	17
Azotés	120-72-9	2.1	8	7	0	0	0	0	0	17	0	0	0	10	1	6	1	0	117	7	6	2	51	90	102	3	13	4	12	3	3	10	266	32	472	9	16	51	26.4	6	18	2	22	25
Azotés	121-69-7	2.3	8	11	0	0	0	0	0	20	3	0	0	6	0	6	0	0	121	6	4	1	65	72	78	3	18	4	17	3	3	9	285	26	528	9	20	56	37.09	6	17	2	25	25
Azotés	583-58-4	1.6	7	9	0	0	0	0	0	17	2	0	0	6	4	1	2	0	107	6	4	1	44	70	75	3	15	4	14	3	3	10	195	23	339	7	17	50	30.11	4	13	2	22	22
Azotés	591-22-0	1.8	7	9	0	0	0	0	0	17	2	0	0	6	4	1	2	0	107	6	4	1	44	70	74	3	15	4	14	3	3	10	198	23	348	7	17	50	30.1	4	13	2	22	22
Azotés	62-53-3	0.9	6	7	0	0	0	0	0	14	1	0	0	6	0	6	0	0	93	7	4	1	29	62	64	2	12	3	11	2	3	9	142	19	240	7	12	35	23.09	4	12	2	18	20
Azotés	83-34-1	2.6	9	9	0	0	0	0	0	20	1	0	0	10	2	5	2	0	131	7	6	2	72	98	113	4	16	5	15	3	3	10	352	36	669	10	20	60	33.23	6	20	2	26	28
Azotés	91-22-5	2	9	7	0	0	0	0	0	18	0	0	0	11	1	7	1	0	129	8	7	2	71	104	118	3	14	5	12	3	3	10	353	37	643	10	20	64	27.73	7	23	2	24	28
Azotés	94-68-8	2.5	9	13	0	0	0	0																																				

Soufrés	463-58-1	1.5	1	0	0	0	0	0	0	1	0	1	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	60	20	2	0	2	24	16	1	3	1	2	1	5	13	6	10	3	1	6	18	6	1	3	4	6	9
Chlorés	503-38-8	3.1	2	0	0	4	0	0	0	2	0	0	0	7	6	1	0	0	0	0	0	0	0	0	0	0	0	198	25	1	0	37	44	40	3	7	3	6	3	6	14	150	34	265	6	28	66	53.5	4	9	3	28	27
Esters	108-65-6	0.4	6	12	0	0	0	0	0	3	0	0	0	20	7	1	0	0	0	4	1	0	0	0	0	0	132	6	1	0	82	44	41	4	19	4	18	3	4	9	300	23	490	7	40	88	51.67	6	13	3	30	27	
Esters	111-15-9	0.3	6	12	0	0	0	0	0	3	0	0	0	20	7	1	0	0	0	6	0	0	0	0	0	132	6	1	0	92	42	38	4	19	4	18	3	3	7	336	23	520	7	41	87	51	6	13	3	30	27		
Soufrés	2365-48-2	0.4	3	6	0	0	0	0	0	2	0	1	0	11	4	1	0	0	0	2	0	0	0	0	0	106	9	1	0	22	30	28	2	11	3	10	2	4	9	83	15	103	4	18	40	37	2	6	3	20	19		
Soufrés	623-51-8	0.8	4	8	0	0	0	0	0	2	0	1	0	14	5	1	0	0	0	4	0	0	0	0	0	120	8	1	0	38	34	32	3	14	3	13	3	4	9	138	18	189	5	25	54	42	3	8	3	24	21		
Soufrés	151-01-9	1.7	3	6	0	0	0	0	0	1	0	2	0	11	4	1	0	0	0	2	0	0	0	0	0	122	10	1	0	21	30	26	2	11	3	11	2	4	9	79	20	94	4	19	43	34.5	2	5	2	20	18		
Aromatiques	104-55-2	1.9	9	8	0	0	0	0	0	1	0	0	0	18	2	2	0	0	6	1	7	1	0	0	0	132	7	6	1	96	92	96	3	15	5	14	3	3	9	411	32	725	10	27	81	27.67	9	27	2	23	28		
Esters	91-64-5	1.4	9	6	0	0	0	0	0	2	0	0	0	18	4	2	0	0	6	1	7	1	0	0	0	146	9	7	2	102	110	124	4	14	5	12	3	3	9	485	41	923	11	24	74	36.17	7	23	3	28	32		
Aromatiques	104-46-1	3.3	10	12	0	0	0	0	0	1	0	0	0	23	4	1	0	0	6	1	7	1	0	0	0	148	6	5	1	124	92	100	4	20	5	19	4	3	9	514	36	966	11	31	85	44.16	9	26	2	31	31		
Aromatiques	523-80-8	2.7	12	14	0	0	0	0	0	4	0	0	0	31	10	1	0	0	6	1	7	1	0	0	0	222	7	6	2	302	128	152	7	26	8	23	5	3	9	1272	67	2504	16	49	130	80.66	11	31	2	50	48		
Aromatiques	94-59-7	3	10	10	0	0	0	0	0	2	0	0	0	23	6	1	0	0	6	1	7	1	0	0	0	162	7	6	2	149	104	117	5	18	6	17	4	2	7	673	43	1310	12	27	78	49.99	7	22	2	34	35		
Esters	119-36-8	2.3	8	8	0	0	0	0	0	3	0	0	0	19	4	1	0	0	6	1	5	1	0	0	0	152	8	5	1	111	94	105	4	16	5	14	3	3	10	465	38	854	10	33	92	44.17	8	26	3	30	33		
Aromatiques	90-05-1	1.3	7	8	0	0	0	0	0	2	0	0	0	17	3	0	0	0	6	0	6	0	0	0	0	124	7	4	1	60	74	81	3	15	4	13	3	3	10	264	28	470	9	21	57	38.67	5	16	3	26	27		
Aldéhydes	120-57-0	1.1	8	6	0	0	0	0	0	3	0	0	0	18	5	1	0	0	6	1	5	1	0	0	0	150	9	6	2	104	100	115	4	14	5	12	3	3	8	489	40	903	11	25	73	44.32	7	20	3	30	33		
Aldéhydes	1205-17-0	2.2	11	12	0	0	0	0	0	3	0	0	0	27	8	1	0	0	6	2	6	2	0	0	0	192	7	6	2	251	114	129	6	22	7	20	5	2	7	1091	52	2168	14	38	105	62.32	9	27	2	41	41		
Aldéhydes	123-11-5	1.8	8	8	0	0	0	0	0	2	0	0	0	18	3	1	0	0	6	1	5	1	0	0	0	136	8	5	1	88	86	94	4	15	5	14	3	3	9	375	33	660	9	28	78	36.99	8	23	3	26	29		
Azotés	134-20-3	1.9	8	9	0	0	0	0	0	1	2	0	0	20	4	1	0	0	6	1	5	1	0	0	0	151	8	5	1	112	94	105	4	17	5	15	3	3	10	469	38	867	10	33	92	43.92	8	26	3	30	32		
Chlorés	98-88-4	2.9	7	5	0	1	0	0	0	1	0	0	0	14	2	1	0	0	6	1	5	1	0	0	0	141	10	5	1	60	82	88	3	12	4	10	3	3	10	270	31	483	9	21	63	31.09	6	20	3	23	26		
Aromatiques	97-53-0	2	10	12	0	0	0	0	0	2	0	0	0	24	5	1	0	0	6	1	7	1	0	0	0	164	7	5	1	154	98	107	4	21	6	19	4	3	9	626	41	1212	12	36	99	49.99	10	28	2	34	35		
Aldéhydes	121-33-5	1.2	8	8	0	0	0	0	0	3	0	0	0	19	4	1	0	0	6	1	5	1	0	0	0	152	8	5	1	109	94	105	4	16	5	14	3	3	10	454	39	832	10	33	92	44.32	8	26	3	30	33		

Annexe 12 Indices de structure des 350 molécules – Partie 2

Famille	N °CAS	Acides	Alcanes	Alcènes	Alcools	Alcynes	Aldehydes	Amides	Amines	Anhydrides	Aromatiques	Bromes	Cétones	Chlores	Esters	Ethers	Fluorés	Iodes	Phosphores	Soufres	Terpenes	Peroxydes	Sulfinyles	CombiAtomeInter
Acides	107-92-6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	109-52-4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	111-14-8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	112-05-0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	124-07-2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	142-62-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	143-07-7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	334-48-5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	503-74-2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	64-18-6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	64-19-7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	646-07-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	79-09-4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Acides	79-31-2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Acides
Alcanes	106-97-8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	108-08-7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	108-87-2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	109-66-0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	110-54-3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	110-82-7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	111-65-9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	111-84-2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	112-40-3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	1120-21-4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	124-18-5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	142-82-5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	292-64-8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	3522-94-9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	540-84-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	563-59-3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	589-34-4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	589-53-7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	589-81-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	590-35-2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	591-76-4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	592-27-8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	617-78-7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	74-98-6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	75-83-2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	78-78-4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	79-29-8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	96-14-0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcanes	96-37-7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcanes
Alcènes	106-98-9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	106-99-0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	109-67-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	110-83-8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	111-66-0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	111-67-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	115-07-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes
Alcènes	115-11-7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcènes

Alcènes	124-11-8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes
Alcènes	592-41-6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes
Alcènes	592-76-7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes
Alcènes	74-85-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes
Alcools	104-76-7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	106-21-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	111-27-3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	111-70-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	111-87-5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	112-30-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	123-51-3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	123-96-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	143-08-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	1490-04-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	1565-80-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	1653-40-3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	19700-21-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	26952-21-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	6032-29-7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	64-17-5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	67-56-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	67-63-0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	71-23-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	71-36-3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	71-41-0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	75-65-0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	75-85-4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	78-83-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	78-92-2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcools	89-78-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcools
Alcynes	74-86-2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcynes
Aldéhydes	110-62-3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	111-71-7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	112-31-2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	123-05-7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	123-38-6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	123-72-8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	124-13-0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	124-19-6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	50-00-0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	590-86-3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	66-25-1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	75-07-0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	78-84-2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Aldéhydes	96-17-3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Aldéhydes
Azotés	107-10-8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	109-73-9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	109-89-7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	110-89-4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	111-26-2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	121-44-8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	123-75-1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	124-40-3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	13952-84-6	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	5724-81-2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines

Azotés	74-89-5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	75-04-7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	75-05-8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	75-31-0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	75-50-3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	75-64-9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	7664-41-7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Azotés	78-81-9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Amines
Aromatiques	100-41-4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	103-65-1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	104-51-8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	105-05-5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	106-42-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	1077-16-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	108-38-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	108-67-8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	108-88-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	119-64-2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	135-01-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	141-93-5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	2189-60-8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	488-23-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	611-14-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	620-14-4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	622-96-8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	71-43-2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	91-20-3	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	91-57-6	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	95-13-6	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	95-47-6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	95-63-6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Aromatiques	98-82-8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques
Bromés	75-25-2	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	Bromes
Cétones	107-87-9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	108-10-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	108-83-8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	108-94-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	110-12-3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	110-43-0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	111-13-7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	112-12-9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	1195-79-5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	120-92-3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	431-03-8	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	563-80-4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	565-61-7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	589-38-8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	591-78-6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	600-14-6	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	67-64-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	693-54-9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	75-97-8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	76-22-2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	78-93-3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones
Cétones	821-55-6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Cétones

Chlorés	56-23-5	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	Chlores
Chlorés	67-66-3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	Chlores
Chlorés	71-55-6	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	Chlores
Chlorés	75-09-2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	Chlores
Chlorés	7647-01-0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Chlores
Esters	105-37-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	105-46-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	105-54-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	105-66-8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	106-27-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	106-30-9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	106-36-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	107-31-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	108-21-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	108-64-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	109-19-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	109-21-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	109-60-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	109-94-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	110-19-0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	110-38-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	110-74-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	112-17-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	123-86-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	123-92-2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	141-06-0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	141-78-6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	142-92-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	32210-23-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	539-82-2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	539-90-2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	540-42-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	540-88-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	542-55-2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	547-63-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	554-12-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	556-24-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	557-00-6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	589-59-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	590-01-2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	592-84-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	617-50-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	623-42-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	624-24-8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	624-54-4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	625-55-8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	627-90-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	628-63-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	637-78-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	638-11-9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	644-49-5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	79-20-9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	97-62-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	97-85-8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters
Esters	97-87-0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Esters

Ethers	109-99-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	Ethers
Ethers	1634-04-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	Ethers
Ethers	3299-32-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	Ethers
Ethers	470-82-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Ethers
Ethers	60-29-7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Ethers
Iodés	75-47-8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	Iodes
Soufrés	107-03-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	109-79-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	110-01-0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	110-66-7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	110-81-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	Soufres
Soufrés	111-31-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	1679-09-0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	352-93-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	513-44-0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	513-53-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	541-31-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	624-92-0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	Soufres
Soufrés	74-93-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	75-08-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	75-15-0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	Soufres
Soufrés	75-18-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	75-33-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	75-66-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	7783-06-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Soufres
Soufrés	7446-09-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Sulfinyles
Soufrés	68-11-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Acides-Soufres
Alcools	106-22-9	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools
Alcools	3391-86-4	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools
Alcools	616-25-1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools
Alcools	78-70-6	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools
Alcools	98-55-5	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools
Aldéhydes	107-02-8	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	123-73-9	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	18829-55-5	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	18829-56-6	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	2463-53-8	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	2548-87-0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	30361-28-5	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	3913-81-3	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	4313-03-5	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	5392-40-5	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	5910-87-2	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	6728-26-3	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	6728-35-4	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Aldéhydes	78-85-3	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes
Azotés	107-13-1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Amines
Azotés	126-98-7	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Amines
Aromatiques	100-42-5	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aromatiques
Cétones	110-93-0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones
Cétones	127-41-3	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones
Cétones	6485-40-1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones
Cétones	78-59-1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones
Cétones	89-82-7	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones
Cétones	99-49-0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Cetones

Chlorés	127-18-4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Chlores
Chlorés	79-01-6	0	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	Alcenes-Chlores
Esters	106-63-8	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Alcenes-Esters
Esters	140-88-5	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Alcenes-Esters
Esters	80-62-6	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Alcenes-Esters
Esters	96-33-3	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Alcenes-Esters
Ethers	1708-29-8	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Alcenes-Ethers
Soufrés	10152-76-8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Soufres
Soufrés	2179-57-9	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	Alcenes-Soufres
Soufrés	592-88-1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	Alcenes-Soufres
Terpènes	127-91-3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	138-86-3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	5989-08-2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	5989-27-5	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	7785-26-4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	7785-70-8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	78-79-5	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	80-56-8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Terpènes	99-85-4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Alcenes-Terpènes
Alcools	10415-87-9	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	106-44-5	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	108-39-4	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	108-95-2	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	60-12-8	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	89-83-8	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Aromatiques	95-48-7	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques
Esters	138-22-7	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Alcools-Esters
Esters	25265-77-4	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Alcools-Esters
Alcools	110-80-5	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Alcools-Ethers
Alcools	111-76-2	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Alcools-Ethers
Ethers	5131-66-8	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Alcools-Ethers
Azotés	104-90-5	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	108-48-5	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	109-06-8	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	110-86-1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	120-72-9	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	121-69-7	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	583-58-4	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	591-22-0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	62-53-3	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	83-34-1	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	91-22-5	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Azotés	94-68-8	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques
Soufrés	556-61-6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	Amines-Soufres
Esters	763-69-9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	Anhydrides-Ethers
Cétones	98-86-2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	Aromatiques-Cétones
Chlorés	108-90-7	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	Aromatiques-Chlores
Aromatiques	122-79-2	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Esters
Esters	140-11-4	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Esters
Esters	93-89-0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Esters
Aromatiques	100-66-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	Aromatiques-Ethers
Aromatiques	93-04-9	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	Aromatiques-Ethers
Ethers	110-00-9	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Ethers
Ethers	271-89-6	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Ethers
Ethers	3777-69-3	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	Aromatiques-Ethers

Soufrés	110-02-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	Aromatiques-Soufres
Chlorés	75-36-5	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Cetones-Chlores
Chlorés	75-44-5	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	Cetones-Chlores
Soufrés	463-58-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	Cetones-Soufres
Chlorés	503-38-8	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	Chlores-Esters
Esters	108-65-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	Esters-Ethers
Esters	111-15-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	Esters-Ethers
Soufrés	2365-48-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	Esters-Soufres
Soufrés	623-51-8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	Esters-Soufres
Soufrés	151-01-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	Ethers-Soufres
Aromatiques	104-55-2	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aldehydes-Aromatiques
Esters	91-64-5	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Aromatiques-Esters
Aromatiques	104-46-1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Alcenes-Aromatiques-Ethers
Aromatiques	523-80-8	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	Alcenes-Aromatiques-Ethers
Aromatiques	94-59-7	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	Alcenes-Aromatiques-Ethers
Esters	119-36-8	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques-Esters
Aromatiques	90-05-1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Alcools-Aromatiques-Ethers
Aldéhydes	120-57-0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	Aldehydes-Aromatiques-Ethers
Aldéhydes	1205-17-0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	Aldehydes-Aromatiques-Ethers
Aldéhydes	123-11-5	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Aldehydes-Aromatiques-Ethers
Azotés	134-20-3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Amines-Aromatiques-Esters
Chlorés	98-88-4	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Aromatiques-Cetones-Chlores
Aromatiques	97-53-0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcenes-Alcools-Aromatiques-Ethers
Aldéhydes	121-33-5	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Alcools-Aldehydes-Aromatiques-Ethers

Annexe 13 Utilisation de l'indice d'Hopkins

La fonction de R utilisée est « hopkins » du package « clustertend » (YILAN et RUTONG 2015) construit à partir de l'article de Lawson (LAWSON et JURIS 1990). Cependant, comme cela est mentionné dans la discussion du site internet (kassambra 2017) son résultat n'est pas cohérent avec la structure du fichier « Iris » (fichier utilisé très régulièrement comme exemple dans R). En effet, la fonction « hopkins » du package « clustertend » calcule l'indicateur « 1-H ». Cette fonction est alors utilisée après rectification de l'indice « 1-H » en « H ».

En effet, certains auteurs semblent travailler avec la grandeur « 1-H » puisqu'ils cherchent à approcher la valeur 0 pour considérer un jeu de données comme « classable » [(J. HAN, KAMBER, et PEI 2012) ; (HERVÉ 2014)]. Récemment, cette ambiguïté est soulevée dans des forums [(kassambra 2017) ; (amwalker8 2018)] concernant les packages R.

Annexe 14 COORDONNEES DES MOLECULES SUR LES AXES PP (ORDRE SELON AXE 1 CROISSANT) 350 MOLECULES

Molécules	Axe 1	Axe 2	Molécules	Axe 1	Axe 2	Molécules	Axe 1	Axe 2
quinoline	-3,53	-0,71	trichloroethene	0,19	-1,03	hydrogen chloride	0,63	-1,83
2-methoxynaphtalene	-3,41	0,26	6-methyl-5-hepten-2-one	0,19	0,59	butyl isovalerate	0,63	1,62
2-methyl naphthalene	-3,24	0,00	(E)-2-heptenal	0,20	0,25	pentyl propanoate	0,63	1,16
naphtalene	-3,23	-0,39	carbon tetrachloride	0,21	-1,19	acide 4-methylpentanoïque	0,63	0,21
skatole	-3,19	-0,57	cyclooctane	0,22	0,13	isopentyl butanoate	0,63	1,62
indole	-3,18	-0,93	(E)-2-pentadiene	0,22	-0,28	butyl 2-hydroxypropanoate	0,63	1,02
2,3-benzofuran	-2,90	-0,64	2-hexenal (E)	0,23	-0,14	methylformate	0,64	-1,12
indene	-2,88	-0,62	trimethylamine	0,24	-1,37	isopropyl propionate	0,64	0,28
apiole	-2,80	1,95	1,3-butadiene	0,26	-1,02	butyl isobutyrate	0,64	1,16
safrole	-2,41	0,56	carbonedisulfide	0,26	-1,55	1-octanol	0,64	0,69
7-methyl-1,3-benzodioxole-5-propanal	-2,35	1,37	cyclohexane	0,26	-0,49	propyl butyrate	0,65	0,71
3,4-(Methylenedioxy) benzaldehyde	-2,33	0,08	carbonyl chloride	0,27	-1,35	acetaldehyde	0,65	-1,33
Coumarin	-2,25	0,19	methylcyclohexane	0,27	-0,17	dodecane	0,65	2,30
methyl 2-aminobenzoate	-2,02	0,16	2,3-butanedione (diacetyl)	0,27	-0,68	acide butanoïque (acide butyrique)	0,65	-0,51
1, 2, 3,4-tetrahydronaphtalene	-2,00	0,01	hexylamine	0,29	-0,35	pentanoate d'ethyl	0,65	0,72
N, N-dimethyl-benzenamine	-1,94	-0,51	methylcyclopentane	0,29	-0,46	acide 3-methylbutanoïque	0,65	-0,15
eugenol	-1,92	0,77	(E)-2-butenal	0,30	-0,83	isopropyl formate	0,65	-0,46
anethole	-1,92	0,52	dimethylamine	0,30	-1,61	3,7-dimethyloctanol	0,65	1,59
3-phenyl-2-propenal	-1,91	0,12	2-methyl-2-propenal	0,31	-0,80	sec, butyl acetate	0,65	0,29
vanillin	-1,84	0,32	triiodomethane	0,31	-0,31	2-ethylhexanol	0,65	0,69
aniline	-1,84	-1,20	Carbonyl sulfide	0,31	-1,52	butyl propanoate	0,65	0,73
N-ethyl-2-methylaniline	-1,80	-0,05	2-butylamine	0,32	-1,03	propyl isovalerate	0,65	1,15
4-methoxybenzaldehyde	-1,80	0,06	ammonia	0,32	-2,10	isobutyl butyrate	0,65	1,15
2,6-dimethylpyridine	-1,79	-0,78	methacrylate de methyl	0,32	-0,14	methylacetate	0,66	-0,78
5-Ethyl-2-methylpyridine	-1,76	-0,39	tert-butylamine	0,33	-1,03	acide 2-methylpropanoïque	0,66	-0,49
2-picoline	-1,72	-1,09	2-undecanone	0,34	2,11	acetate de 1,1-dimethylethyle	0,66	0,29
benzoyl chloride	-1,71	-0,33	butylamine	0,34	-0,99	Hexyl acetate	0,66	1,17
styrene	-1,70	-0,48	methyl acrylate	0,35	-0,50	isobutyl isovalerate	0,66	1,62
3,4-dimethylpyridine	-1,68	-0,70	1-nonene	0,35	0,81	propyl isobutyrate	0,66	0,73
3,5-dimethylpyridine	-1,67	-0,69	2-propenal (acrolein)	0,35	-1,13	n-butyl formate	0,66	-0,08
acetophenone	-1,67	-0,25	isopropylamine	0,36	-1,31	1-butoxy-2-propanol	0,67	0,60
pyridine	-1,66	-1,38	methylamine	0,36	-1,85	1-heptanol	0,67	0,30
2-methoxyphenol	-1,63	-0,35	2-decanone	0,36	1,59	isobutyl isobutyrate	0,67	1,17
octylbenzene	-1,62	2,27	1-octen-3-ol	0,37	0,62	ethylbutyrate	0,67	0,31
methyl salicylate	-1,61	0,45	isobutylamine	0,37	-0,99	n-propyl propionate	0,67	0,33
methoxybenzene	-1,60	-0,57	2-octene	0,38	0,34	acide propanoïque	0,68	-0,83
hexylbenzene	-1,55	1,22	propylamine	0,38	-1,27	undecane	0,68	1,78
3-methyl-1-phenyl-3-pentanol	-1,54	1,47	1-octene	0,38	0,39	ethyl-3-methylbutanoate	0,68	0,72
chlorobenzene	-1,53	-0,91	decanal	0,39	1,64	6-Methyl-1-heptanol	0,68	0,70
o-diethylbenzene	-1,52	0,29	A-citronellol	0,39	1,53	isooctanol	0,68	0,70
m-diethylbenzene	-1,51	0,30	nonan-2-one	0,39	1,10	isobutylpropionate	0,68	0,74
1,4-diethylbenzene	-1,51	0,30	chloroforme	0,40	-1,28	pentyl acetate (=amyl acetate)	0,68	0,72
2-isopropenyl-5-methylphenol	-1,50	0,58	tetrahydrothiophene	0,40	-0,81	ethyl isobutyrate	0,68	0,33
butylbenzene	-1,50	0,33	1-heptene	0,41	0,01	isopropyl acetate	0,69	-0,09
ethyl benzoate	-1,49	0,64	nonanal	0,42	1,15	n-propyl formate	0,69	-0,43
2-ethyltoluene	-1,48	-0,07	isobutyl-2-propanoate	0,42	0,73	isobutyl formate	0,69	-0,07
1, 2, 3,4-tetramethylbenzene	-1,48	0,30	2-octanone	0,42	0,64	ethyl propionate	0,70	-0,04
propylbenzene	-1,48	-0,06	ethylamine	0,42	-1,54	2-methyl-2-butanol	0,70	-0,45
1-ethyl-3-methylbenzene	-1,47	-0,07	2,6-dimethyl-4-heptanone	0,43	1,07	hexanol	0,70	-0,08
p-ethyltoluene	-1,47	-0,06	acide n-dodecanoïque	0,43	3,27	2-pentanol	0,70	-0,45
phenyl acetate	-1,47	0,15	ethyl acrylate	0,44	-0,07	acetate de butyle	0,70	0,33

1-methylethylbenzene	-1,46	-0,05	1, 1,1-trichloroethane	0,44	-1,00	decane	0,70	1,29
ethylbenzene	-1,46	-0,43	2-ethylhexanal	0,44	0,69	acide formique	0,71	-1,44
2-methylphenol	-1,45	-0,56	1-hexene	0,44	-0,33	3-methylbutyl acetate	0,71	0,73
1, 2,4-trimethylbenzene	-1,45	-0,08	Allyl sulfide	0,44	-0,06	Formate d'ethyle	0,72	-0,75
1, 3,5-trimethylbenzene	-1,45	-0,07	3-hexanone	0,45	-0,15	2-butoxyethanol	0,72	0,24
3-methylphenol	-1,45	-0,56	octanal	0,45	0,70	acide acetique	0,72	-1,13
o-xylene	-1,45	-0,42	2-heptanone	0,45	0,24	n-propyl acetate	0,72	-0,05
4-methylphenol (p. cresol)	-1,44	-0,56	1-pentene-3-ol	0,45	-0,50	1-pentanol	0,73	-0,41
m-xylene	-1,44	-0,42	Ethanamine, N-ethyl-	0,46	-0,92	2-ethoxyethyl acetate	0,73	0,68
p-xylene	-1,43	-0,42	tribromomethane	0,46	-0,78	acetate d'isobutyl	0,73	0,34
phenol	-1,43	-0,90	triethylamine	0,46	-0,24	2-butanol	0,73	-0,75
toluene	-1,43	-0,76	3,3-dimethyl-2-butanone	0,46	-0,07	nonane	0,73	0,85
acetate de benzyle	-1,42	0,62	3-methyl-2-pentanone	0,47	-0,13	methyl 2-mercaptoacetate	0,74	-0,55
2-phenylethanol	-1,42	-0,20	2-hexanone	0,47	-0,14	diethylether	0,74	-0,64
benzene	-1,41	-1,10	1-pentene	0,47	-0,67	2-methyl-2-propanol	0,74	-0,75
(1R,2S,7R,8R)-2,6,6,9-tetramethyltricyclo[5,4,0,0(2,8)]undec-9-ene	-1,29	2,49	heptanal	0,48	0,29	acetate d'ethyle	0,74	-0,39
2-pentylfuran	-1,16	0,42	5-methyl-2-hexanone	0,48	0,24	(S)-2-methyl-1-Butanol	0,75	-0,41
furane	-1,04	-1,26	Acide decanoique	0,48	2,10	octane	0,76	0,45
tiophene	-0,85	-1,25	ethyl undecanoate	0,49	4,03	2-methyl-2-butanethiol	0,76	-0,45
methacrylonitrile	-0,79	-1,31	ethanoyl chloride	0,50	-1,18	3-methyl-1-butanol	0,76	-0,41
acrylonitrile	-0,74	-1,66	2-pentanone	0,50	-0,49	1-butanol	0,76	-0,72
(1S,5S)-2,6,6-Trimethylbicyclo[3,1,1]hept-2-ene	-0,69	0,63	4-methyl-2-pentanone	0,50	-0,13	Hydrogen sulfide	0,76	-1,83
(1R,5R)-2,6,6-Trimethylbicyclo[3,1,1]hept-2-ene	-0,69	0,63	1-butene	0,50	-0,94	4-methylheptane	0,77	0,47
alpha-pinene	-0,69	0,63	hexanal	0,50	-0,09	2-Ethoxyethanol	0,77	-0,46
Beta-pinene	-0,69	0,64	3-methyl-2-butanone	0,50	-0,47	isopropanol	0,78	-1,03
4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one	-0,66	2,29	Sulfure dioxide	0,51	-1,67	3-methylheptane	0,78	0,47
camphre	-0,62	0,89	Acide nonanoique	0,52	1,57	heptane	0,78	0,12
Fenchone	-0,62	0,90	2-methyl-1-propene	0,52	-0,93	1-hexanethiol	0,78	-0,04
(R)-5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-0,58	0,93	ethyl decanoate	0,52	3,36	3-ethylpentane	0,79	0,10
5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-0,58	0,93	2-butanone	0,52	-0,81	2-methylpropanol	0,79	-0,71
Geosmine	-0,50	1,60	pentanal	0,54	-0,45	2-methyl-Heptane	0,79	0,47
eucalyptol	-0,50	0,79	2-methylbutanal	0,54	-0,43	2-butanethiol	0,79	-0,74
gamma-terpinene	-0,45	0,71	Acide octanoique	0,54	1,08	mercaptoacetic acid	0,80	-0,91
Limonene	-0,42	0,73	decyl acetate	0,55	3,37	methanol	0,80	-1,57
R-Limonene	-0,42	0,73	1-propene	0,55	-1,23	tert-butanethiol	0,80	-0,75
acetonitrile	-0,42	-1,90	dichloromethane	0,55	-1,39	1-propanol	0,80	-1,01
(5R)-5-methyl-2-propan-2-ylidene-cyclohexan-1-one	-0,37	0,97	Diallyl disulfide	0,55	0,28	3-methylhexane	0,80	0,10
3, 3,5-trimethyl-2-cyclohex-1-ene	-0,32	0,63	texanol	0,56	3,68	hexane	0,80	-0,22
1-pyrroline	-0,31	-1,24	ethyl-3-ethoxy-propionate	0,56	1,02	1-pentanethiol	0,81	-0,40
acetylene	-0,19	-2,02	3-methylbutanal	0,56	-0,44	dimethylsulfide(DMS)	0,81	-1,17
(E, E)-2,4-nonadienal	-0,15	1,09	butanal	0,56	-0,75	2, 2, 5-trimethylhexane	0,81	0,93
alpha-terpineol	-0,13	0,96	methyl tert-butyl ether	0,57	-0,46	2-methylhexane	0,81	0,11
methyl isothiocyanate	-0,13	-1,54	1-methoxy-2-propyl acetate	0,57	0,56	2, 2-dimethylpentane	0,82	0,17
piperidine	-0,12	-0,88	acetone	0,57	-1,10	isopropyl mercaptane	0,82	-1,01
3,7-dimethyl-2,6-octadienal	-0,12	1,54	methyl valerate	0,57	0,27	2,3-dimethyl pentane	0,83	0,11
(E, E)-2,4-octadienal	-0,11	0,63	ethyl ethoxydithioformate	0,57	-0,66	butanethiol	0,83	-0,68
cyclohexene	-0,11	-0,61	Acide heptanoique	0,57	0,63	3-methylpentane	0,83	-0,23
pyrrolidine	-0,10	-1,15	2-methylpropanal	0,58	-0,74	methanethiol	0,83	-1,49
(E, E)-2,4-heptadienal	-0,08	0,20	1-decanol	0,59	1,64	isooctane	0,83	0,53
cyclohexanone	-0,06	-0,46	methylbutyrate	0,59	-0,11	ethyl-2-mercaptoacetate	0,83	-0,13
2-5-dihydrofuran	-0,05	-0,94	methyl 3-methylbutanoate	0,60	0,27	pentane	0,83	-0,52
cyclopentanone	-0,04	-0,76	methyl isobutyrate	0,60	-0,10	Isoamyl mercaptane	0,83	-0,38
tetrachloroethylene	0,03	-0,86	propanal	0,60	-1,04	2,2-dimethylbutane	0,84	-0,16
3,7-dimethylocta-1,6-dien-3-ol	0,07	1,46	acide hexanoique	0,60	0,22	2,4-dimethyl pentane	0,85	0,11
2, 4,5-trimethyl-1,3-dioxolane	0,07	-0,05	ethyl heptanoate	0,60	1,64	ethanol	0,85	-1,26
4-tert-butylcyclohexylacetate	0,09	2,50	formaldehyde	0,61	-1,60	n-propyl mercaptane	0,85	-0,96

(E)-2-decenal	0,11	1,63	ethylene	0,61	-1,46	1-isobutanethiol	0,86	-0,68
menthol (89-78-1)	0,12	1,06	methyl propionate	0,61	-0,46	2, 3-dimethylbutane	0,86	-0,20
menthol	0,12	1,06	nonan-1-ol	0,61	1,16	butane	0,87	-0,80
(E)-non-2-enal	0,14	1,13	isopropyl butanoate	0,62	0,67	2-methylbutane	0,87	-0,53
2-nonenal	0,14	1,13	2-octanol	0,62	0,66	dimethyldisulfide (DMS)	0,88	-0,90
isoprene	0,15	-0,66	propyl valerate	0,62	1,15	Ethanethiol	0,89	-1,22
6-nonenal	0,16	1,08	butylbutyrate	0,62	1,14	Propane	0,91	-1,07
(E)-2-octenal	0,17	0,67	acide pentanoïque	0,62	-0,16	Diethyl sulfide	0,94	-0,57
trichloromethyl chloroformate	0,18	-0,33	Methyl allyl sulfide	0,63	-0,68	Diethyl disulfide	1,05	-0,30
tetrahydrofurane	0,19	-0,88	isopropyl isobutyrate	0,63	0,70			

Annexe 15 COORDONNEES DES INDICES DE STRUCTURE SUR LES AXES PP (ORDRE SELON AXE 1 CROISSANT) (47 INDICES)

Indices	Axe 1	Axe 2
Ni	-0,95	0,08
NbLConj	-0,93	-0,06
Aromatiques	-0,93	-0,07
M2	-0,85	0,46
CYCLE	-0,85	0,09
M1	-0,84	0,48
Jt...CEP.	-0,55	0,75
C	-0,44	0,85
X1X	-0,43	0,89
Jt.Dh.	-0,36	0,00
J...CEP.	-0,34	0,87
Ethers.	-0,29	0,08
GUTMAN	-0,27	0,89
MM_Molecule	-0,26	0,79
N	-0,25	-0,34
Amines	-0,25	-0,34
MTI	-0,24	0,91
XLogP3	-0,20	0,64
X1Xv	-0,18	0,90
W.Dh.	-0,16	0,93
Jt..Ah.	-0,14	0,92
Terpenes	-0,09	0,12
Alcenes	-0,08	0,14
NbLT	-0,06	-0,18
nQuaternaire	-0,05	0,25
Aldehydes	-0,01	0,06
Alcynes	-0,01	-0,11
X0X	0,01	0,93
Cetones	0,02	0,03
Br	0,02	-0,04
Bromes	0,02	-0,04
Cl	0,02	-0,17
Chlores	0,02	-0,17
Sulfinyles	0,03	-0,09
J..Ah.	0,04	0,93
W...CEP.	0,04	0,91
X0Xv	0,05	0,92
J.Dh.	0,07	0,01
NbLD	0,10	0,27
Acides	0,13	0,05
O	0,13	0,39
SiTot	0,15	-0,18
Soufres	0,19	-0,22
S	0,19	-0,23
Esters	0,20	0,30
Alcanes	0,22	0,06
nCprimaire	0,36	0,49

**Annexe 16 RESULTAT DE LA CLASSIFICATION DES MOLECULES A PARTIR DE LEURS
COORDONNEES SUR LES DEUX AXES PRINCIPAUX DE LA « PROJECTION POURSUIT »**

Famille	CAS	Molécules	Axe 1	Axe 2	SP	Ordre de grandeur	Classes
Azotés	91-22-5	quinoline	-3,33	-0,71	3E-02	-2	1
Aromatiques	93-04-9	2-méthoxynaphtalène	-3,41	0,26	1E-03	-3	1
Aromatiques	91-57-6	2-méthyl naphtalène	-3,24	0,00	4E-03	-3	1
Aromatiques	91-20-3	naphtalène	-3,23	-0,39	7E-03	-3	1
Azotés	83-34-1	skatole	-3,19	-0,57	4E-04	-4	1
Azotés	120-72-9	indole	-3,18	-0,93	1E-03	-3	1
Ethers	271-89-6	2,3-benzofuran	-2,90	-0,64	5E-02	-2	1
Aromatiques	95-13-6	indène	-2,88	-0,62	2E-02	-2	1
Aromatiques	94-59-7	safrôle	-2,41	0,56	6E-03	-3	1
Aldéhydes	120-57-0	3,4-(Méthylendioxy) benzaldéhyde	-2,33	0,08	7E-03	-3	1
Esters	91-64-5	Coumarin	-2,25	0,19	1E-02	-2	1
Azotés	134-20-3	méthyl 2-aminobenzoate	-2,02	0,16	5E-03	-3	1
Aromatiques	119-64-2	1, 2, 3,4-tetrahydronaphtalène	-2,00	0,01	5E-02	-2	1
Azotés	121-69-7	N, N-diméthyl-benzenamine	-1,94	-0,51	8E-03	-3	1
Aromatiques	97-53-0	eugénol	-1,92	0,77	2E-01	-1	1
Aromatiques	104-46-1	anethole	-1,92	0,52	6E-02	-2	1
Aromatiques	104-55-2	3-phényl-2-propenal	-1,91	0,12	2E-02	-2	1
Aldéhydes	121-33-5	vanilline	-1,84	0,32	1E-04	-4	1
Azotés	62-53-3	aniline	-1,84	-1,20	2E+00	0	1
Azotés	94-68-8	N-éthyl-2-méthylaniline	-1,80	-0,05	1E-01	-1	1
Aldéhydes	123-11-5	4-méthoxybenzaléhyde	-1,80	0,06	3E-02	-2	1
Azotés	108-48-5	2,6-diméthylpyridine	-1,79	-0,78	3E-03	-3	1
Azotés	104-90-5	5-Éthyl-2-méthylpyridine	-1,76	-0,39	4E-03	-3	1
Azotés	109-06-8	2-picoline	-1,72	-1,09	1E-02	-2	1
Chlorés	98-88-4	benzoyl chloride	-1,71	-0,33	1E-02	-2	1
Aromatiques	100-42-5	styrène	-1,70	-0,48	6E-02	-2	1
Azotés	583-58-4	3,4-diméthylpyridine	-1,68	-0,70	5E-03	-3	1
Azotés	591-22-0	3,5-diméthylpyridine	-1,67	-0,69	5E-03	-3	1
Cétones	98-86-2	acetophenone	-1,67	-0,25	1E-02	-2	1
Azotés	110-86-1	pyridine	-1,66	-1,38	8E-02	-2	1
Aromatiques	90-05-1	2-méthoxyphénol	-1,63	-0,35	3E-03	-3	1
Esters	119-36-8	méthyl salicylate	-1,61	0,45	5E-02	-2	1
Aromatiques	100-66-3	méthoxybenzène	-1,60	-0,57	6E-01	-1	1
Chlorés	108-90-7	chlorobenzène	-1,53	-0,91	1E+00	0	1
Aromatiques	135-01-3	o-diéthylbenzène	-1,52	0,29	5E-02	-2	1
Aromatiques	141-93-5	m-diéthylbenzène	-1,51	0,30	4E-01	-1	1
Aromatiques	105-05-5	1,4-diéthylbenzène	-1,51	0,30	2E-03	-3	1
Aromatiques	89-83-8	2-isopropényl-5-méthylphénol	-1,50	0,58	4E-02	-2	1
Aromatiques	104-51-8	butylbenzène	-1,50	0,33	3E-02	-2	1
Esters	93-89-0	éthyl benzoate	-1,49	0,64	5E-02	-2	1
Aromatiques	611-14-3	2-éthyltoluène	-1,48	-0,07	4E-01	-1	1
Aromatiques	488-23-3	1, 2, 3,4-tetraméthylbenzène	-1,48	0,30	6E-02	-2	1
Aromatiques	103-65-1	propylbenzène	-1,48	-0,06	2E-02	-2	1
Aromatiques	620-14-4	1-éthyl-3-méthylbenzène	-1,47	-0,07	9E-02	-2	1
Aromatiques	622-96-8	p-éthyltoluène	-1,47	-0,06	4E-02	-2	1
Aromatiques	122-79-2	phényl acetate	-1,47	0,15	2E-01	-1	1
Aromatiques	98-82-8	1-méthylethylbenzène	-1,46	-0,05	6E-02	-2	1
Aromatiques	100-41-4	éthylbenzène	-1,46	-0,43	1E-01	-1	1
Aromatiques	95-48-7	2-méthylphénol	-1,45	-0,56	1E-03	-3	1
Aromatiques	95-63-6	1,2,4-triméthylbenzène	-1,45	-0,08	6E-01	-1	1
Aromatiques	108-67-8	1, 3,5-triméthylbenzène	-1,45	-0,07	2E-01	-1	1
Aromatiques	108-39-4	3-méthylphénol	-1,45	-0,56	5E-04	-4	1
Aromatiques	95-47-6	o-xylène	-1,45	-0,42	1E+00	0	1
Aromatiques	106-44-5	4-méthylphénol (p.cresol)	-1,44	-0,56	2E-04	-4	1
Aromatiques	108-38-3	m-xylène	-1,44	-0,42	5E-01	-1	1
Aromatiques	106-42-3	p-xylène	-1,43	-0,42	4E-01	-1	1
Aromatiques	108-95-2	phénol	-1,43	-0,90	5E-02	-2	1
Aromatiques	108-88-3	toluène	-1,43	-0,76	1E+00	0	1
Esters	140-11-4	acetate de benzyle	-1,42	0,62	2E-01	-1	1
Aromatiques	60-12-8	2-phényléthanol	-1,42	-0,20	2E-02	-2	1
Aromatiques	71-43-2	benzène	-1,41	-1,10	9E+00	0	1
Ethers	3777-69-3	2-pentylfuran	-1,16	0,42	3E-01	-1	1
Aromatiques	523-80-8	apiolé	-2,80	1,95	6E+02	2	2
Aldéhydes	1205-17-0	?-méthyl-1,3-benzodioxole-5-propanal	-2,35	1,37	1E-03	-3	2
Aromatiques	2189-60-8	octylbenzène	-1,62	2,27	7E-01	-1	2
Aromatiques	1077-16-3	hexylbenzène	-1,55	1,22	3E-02	-2	2
Alcools	10415-87-9	3-méthyl-1-phényl-3-pentanol	-1,54	1,47	2E-01	-1	2
Terpènes	5989-08-2	(1R,2S,7R,8R)-2,6,6,9-tetraméthyltricyclo[5.4.0.0(2.8)]undéc-9-ène	-1,29	2,49	9E-01	-1	2
Cétones	127-41-3	4-(2,6,6-triméthyl-2-cyclohexen-1-yl)-3-buten-2-one	-0,66	2,29	5E-05	-5	2
Aldéhydes	4313-03-5	(E, E)-2,4-heptadienal	-0,08	0,20	6E-02	-2	3
Ethers	3299-32-9	2, 4,5-triméthyl-1,3-dioxolane	0,07	-0,05	4E-01	-1	3
Chlorés	503-38-8	trichlorométhyl chloroformate	0,18	-0,33	8E-03	-3	3
Aldéhydes	18829-55-5	(E)-2-heptenal	0,20	0,25	2E-01	-1	3
Alcane	292-64-8	cyclooctane	0,22	0,13	4E+00	0	3
Cétones	600-14-6	2,3-pentadione	0,22	-0,28	6E-02	-2	3
Aldéhydes	6728-26-3	2-hexenal (E)	0,23	-0,14	5E-01	-1	3
Alcane	110-82-7	cyclohexane	0,26	-0,49	4E+01	1	3
Alcane	108-87-2	méthylcyclohexane	0,27	-0,17	6E-01	-1	3

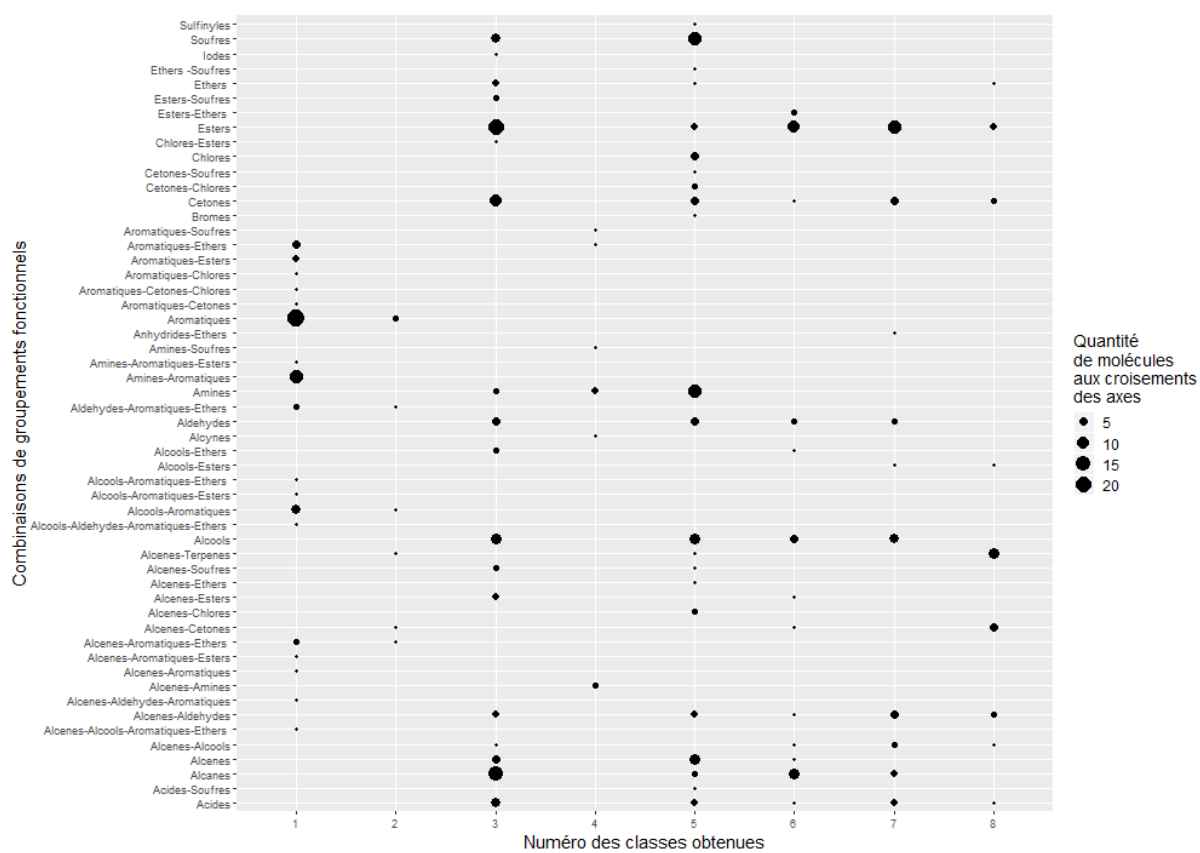
Azotés	111-26-2	hexylamine	0,29	-0,35	6E-02	-2	3
Alcane	96-37-7	methylcyclopentane	0,29	-0,46	6E+00	0	3
Iodés	75-47-8	triiodomethane	0,31	-0,31	6E-02	-2	3
Esters	80-62-6	methacrylate de methyl	0,32	-0,14	7E-01	-1	3
Esters	96-33-3	methyl acrylate	0,35	-0,50	1E-02	-2	3
Alcènes	111-67-1	2-octene	0,38	0,34	2E+00	0	3
Alcènes	111-66-0	1-octene	0,38	0,39	7E-03	-3	3
Alcènes	592-76-7	1-heptene	0,41	0,01	1E-01	-1	3
Esters	140-88-5	ethyl acrylate	0,44	-0,07	1E-03	-3	3
Alcènes	592-41-6	1-hexene	0,44	-0,33	5E-04	-4	3
Soufrés	592-88-1	Allyl sulfide	0,44	-0,06	1E-03	-3	3
Cétones	589-38-8	3-hexanone	0,45	-0,15	1E-01	-1	3
Cétones	110-43-0	2-heptanone	0,45	0,24	8E-02	-2	3
Alcools	616-25-1	1-pentene-3-ol	0,45	-0,50	4E+00	0	3
Azotés	121-44-8	triethylamine	0,46	-0,24	4E-02	-2	3
Cétones	75-97-8	3,3-dimethyl-2-butanone	0,46	-0,07	2E-01	-1	3
Cétones	565-61-7	3-methyl-2-pentanone	0,47	-0,13	1E-01	-1	3
Cétones	591-78-6	2-hexanone	0,47	-0,14	6E-01	-1	3
Aldéhydes	111-71-7	heptanal	0,48	0,29	2E-02	-2	3
Cétones	110-12-3	5-methyl-2-hexanone	0,48	0,24	1E-02	-2	3
Cétones	107-87-9	2-pentanone	0,50	-0,49	2E+00	0	3
Cétones	108-10-1	4-methyl-2-pentanone	0,50	-0,13	7E-01	-1	3
Aldéhydes	66-25-1	hexanal	0,50	-0,09	3E-03	-3	3
Cétones	563-80-4	3-methyl-2-butanone	0,50	-0,47	2E+00	0	3
Aldéhydes	110-62-3	pentanal	0,54	-0,45	1E-02	-2	3
Aldéhydes	96-17-3	2-methylbutanal	0,54	-0,43	2E-03	-3	3
Soufrés	2179-57-9	Diallyl disulfide	0,55	0,28	1E-03	-3	3
Aldéhydes	590-86-3	3-methylbutanal	0,56	-0,44	2E-04	-4	3
Ethers	1634-04-4	methyl tert-butyl ether	0,57	-0,46	2E-02	-2	3
Esters	624-24-8	methyl valerate	0,57	0,27	1E-02	-2	3
Esters	623-42-7	methylbutyrate	0,59	-0,11	3E-02	-2	3
Esters	556-24-1	methyl 3-methylbutanoate	0,60	0,27	1E-02	-2	3
Esters	547-63-7	methyl isobutyrate	0,60	-0,10	8E-03	-3	3
Acides	142-62-1	acide hexanoïque	0,60	0,22	6E-03	-3	3
Esters	554-12-1	methyl propionate	0,61	-0,46	3E-01	-1	3
Acides	109-52-4	acide pentanoïque	0,62	-0,16	1E-03	-3	3
Acides	646-07-1	acide 4-méthylpentanoïque	0,63	0,21	2E-03	-3	3
Esters	637-78-5	isopropyl propionate	0,64	0,28	2E-02	-2	3
Acides	107-92-6	acide butanoïque (acide butyrique)	0,65	-0,51	5E-04	-4	3
Acides	503-74-2	acide 3-méthylbutanoïque	0,65	-0,15	3E-04	-4	3
Esters	625-55-8	isopropyl formate	0,65	-0,46	1E+00	0	3
Esters	105-46-4	sec-butyl acetate	0,65	0,29	1E-02	-2	3
Acides	79-31-2	acide 2-méthylpropanoïque	0,66	-0,49	5E-03	-3	3
Esters	540-88-5	acetate de 1,1-diméthylethyle	0,66	0,29	3E-01	-1	3
Esters	592-84-7	n-butyl formate	0,66	-0,08	4E-01	-1	3
Alcools	111-70-6	1-heptanol	0,67	0,30	1E-01	-1	3
Esters	105-54-4	ethylbutyrate	0,67	0,31	6E-04	-4	3
Esters	106-36-5	n-propyl propionate	0,67	0,33	3E-01	-1	3
Esters	97-62-1	ethyl isobutyrate	0,68	0,33	1E-04	-4	3
Esters	108-21-4	isopropyl acetate	0,69	-0,09	1E+00	0	3
Esters	110-74-7	n-propyl formate	0,69	-0,43	4E+00	0	3
Esters	542-55-2	isobutyl formate	0,69	-0,07	2E+00	0	3
Esters	105-37-3	ethyl propionate	0,70	-0,04	2E-02	-2	3
Alcools	75-85-4	2-methyl-2-butanol	0,70	-0,45	3E+00	0	3
Alcools	111-27-3	hexanol	0,70	-0,08	3E-02	-2	3
Alcools	6032-29-7	2-pentanol	0,70	-0,45	1E+00	0	3
Esters	123-86-4	acetate de butyle	0,70	0,33	5E-02	-2	3
Alcools	111-76-2	2-butoxyethanol	0,72	0,24	2E-01	-1	3
Esters	109-60-4	n-propyl acetate	0,72	-0,05	6E-01	-1	3
Alcools	71-41-0	1-pentanol	0,73	-0,41	4E-01	-1	3
Esters	110-19-0	acetate d'isobutyl	0,73	0,34	4E-02	-2	3
Soufrés	2365-48-2	methyl 2-mercaptoacetate	0,74	-0,55	1E-03	-3	3
Ethers	60-29-7	diethylether	0,74	-0,64	1E-01	-1	3
Esters	141-78-6	acetate d'ethyle	0,74	-0,39	3E+00	0	3
Alcools	1565-80-6	(S)-2-méthyl-1-Butanol	0,75	-0,41	6E-01	-1	3
Soufrés	1679-09-0	2-méthyl-2-butanethiol	0,76	-0,45	2E-03	-3	3
Alcools	123-51-3	3-méthyl-1-butanol	0,76	-0,41	9E-02	-2	3
Alcools	110-80-5	2-Ethoxyethanol	0,77	-0,46	2E+00	0	3
Alcane	142-82-5	heptane	0,78	0,12	3E+00	0	3
Soufrés	111-31-9	1-hexanethiol	0,78	-0,04	7E-05	-5	3
Alcane	617-78-7	3-ethylpentane	0,79	0,10	2E+00	0	3
Alcane	589-34-4	3-methylhexane	0,80	0,10	3E+00	0	3
Alcane	110-54-3	hexane	0,80	-0,22	1E+01	1	3
Soufrés	110-66-7	1-pentanethiol	0,81	-0,40	3E-06	-6	3
Alcane	591-76-4	2-methylhexane	0,81	0,11	2E+00	0	3
Alcane	590-35-2	2, 2-diméthylpentane	0,82	0,17	2E+02	2	3
Alcane	565-59-3	2,3-diméthyl pentane	0,83	0,11	2E+01	1	3
Alcane	96-14-0	3-méthylpentane	0,83	-0,23	3E+01	1	3
Soufrés	623-51-8	ethyl-2-mercaptoacetate	0,83	-0,13	8E-03	-3	3
Alcane	109-66-0	pentane	0,83	-0,52	4E+00	0	3
Soufrés	541-31-1	Isoamyl mercaptane	0,83	-0,38	3E-06	-6	3
Alcane	75-83-2	2,2-diméthylbutane	0,84	-0,16	7E+01	1	3
Alcane	108-08-7	2,4-diméthyl pentane	0,85	0,11	4E+00	0	3
Alcane	79-29-8	2, 3-diméthylbutane	0,86	-0,20	2E+00	0	3

Alcenes	78-78-4	2-methylbutane	0,87	-0,53	4E+00	0	3
Soufrés	352-93-2	Diethyl sulfide	0,94	-0,57	8E-04	-4	3
Soufrés	110-81-6	Diethyl disulfide	1,05	-0,30	2E-03	-3	3
Ethers	110-00-9	furane	-1,04	-1,26	3E+01	1	4
Soufrés	110-02-1	tiophene	-0,85	-1,25	3E-03	-3	4
Azotés	126-98-7	methacrylonitrile	-0,79	-1,31	8E+00	0	4
Azotés	107-13-1	acrylonitrile	-0,74	-1,66	2E+01	1	4
Azotés	75-05-8	acetonitrile	-0,42	-1,90	2E+01	1	4
Azotés	5724-81-2	1-pyrroline	-0,31	-1,24	8E-02	-2	4
Alcynes	74-86-2	acetylene	-0,19	-2,02	2E+02	2	4
Soufrés	556-61-6	methyl isothiocyanate	-0,13	-1,54	5E+00	0	4
Azotés	123-75-1	pyrrolidine	-0,10	-1,15	1E+00	0	4
Azotés	110-89-4	piperidine	-0,12	-0,88	5E-01	-1	5
Alcènes	110-83-8	cyclohexene	-0,11	-0,61	2E-01	-1	5
Cétones	108-94-1	cyclohexanone	-0,06	-0,46	1E+00	0	5
Ethers	1708-29-8	2-5-dihydrofuran	-0,05	-0,94	3E-01	-1	5
Cétones	120-92-3	cyclopentanone	-0,04	-0,76	9E-01	-1	5
Chlorés	127-18-4	tetrachloroethylene	0,03	-0,86	1E+01	1	5
Terpènes	78-79-5	isoprene	0,15	-0,66	1E-01	-1	5
Ethers	109-99-9	tetrahydrofurane	0,19	-0,88	2E+00	0	5
Chlorés	79-01-6	trichloroethene	0,19	-1,03	3E+00	0	5
Chlorés	56-23-5	carbon tetrachloride	0,21	-1,19	3E+01	1	5
Azotés	75-50-3	trimethylamine	0,24	-1,37	1E-04	-4	5
Alcènes	106-99-0	1,3-butadiene	0,26	-1,02	1E+00	0	5
Soufrés	75-15-0	carbonedisulfide	0,26	-1,55	1E-01	-1	5
Chlorés	75-44-5	carbonyl chloride	0,27	-1,35	5E-01	-1	5
Cétones	431-03-8	2,3-butanedione (diacetyl)	0,27	-0,68	1E-04	-4	5
Aldéhydes	123-73-9	(E)-2-butenal	0,30	-0,83	3E-01	-1	5
Azotés	124-40-3	dimethylamine	0,30	-1,61	3E-02	-2	5
Aldéhydes	78-85-3	2-methyl-2-propenal	0,31	-0,80	2E-02	-2	5
Soufrés	463-58-1	Carbonyl sulfide	0,31	-1,52	1E-01	-1	5
Azotés	13952-84-6	2-butylamine	0,32	-1,03	5E-01	-1	5
Azotés	7664-41-7	ammonia	0,32	-2,10	1E+00	0	5
Azotés	75-64-9	tert-butylamine	0,33	-1,03	5E-01	-1	5
Azotés	109-73-9	butylamine	0,34	-0,99	5E-01	-1	5
Aldéhydes	107-02-8	2-propenal (acrolein)	0,35	-1,13	2E-02	-2	5
Azotés	75-31-0	isopropylamine	0,36	-1,31	6E-02	-2	5
Azotés	74-89-5	methylamine	0,36	-1,85	7E-03	-3	5
Azotés	78-81-9	isobutylamine	0,37	-0,99	4E-03	-3	5
Azotés	107-10-8	propylamine	0,38	-1,27	1E-01	-1	5
Chlorés	67-66-3	chloroforme	0,40	-1,28	3E+01	1	5
Soufrés	110-01-0	tetrahydrothiophene	0,40	-0,81	2E-03	-3	5
Azotés	75-04-7	ethylamine	0,42	-1,54	8E-02	-2	5
Chlorés	71-55-6	1, 1,1-trichloroethane	0,44	-1,00	5E+00	0	5
Azotés	109-89-7	Ethanamine, N-ethyl-	0,46	-0,92	1E-01	-1	5
Bromés	75-25-2	tribromomethane	0,46	-0,78	2E+00	0	5
Alcènes	109-67-1	1-pentene	0,47	-0,67	3E-01	-1	5
Chlorés	75-36-5	ethanoyl chloride	0,50	-1,18	1E-01	-1	5
Alcènes	106-98-9	1-butene	0,50	-0,94	8E-01	-1	5
Soufrés	7446-09-5	Sulfure dioxide	0,51	-1,67	2E+00	0	5
Alcènes	115-11-7	2-methyl-1-propene	0,52	-0,93	2E+01	1	5
Cétones	78-93-3	2-butanone	0,52	-0,81	3E+00	0	5
Alcènes	115-07-1	1-propene	0,55	-1,23	2E+01	1	5
Chlorés	75-09-2	dichloromethane	0,55	-1,39	6E+02	2	5
Aldéhydes	123-72-8	butanal	0,56	-0,75	1E-03	-3	5
Cétones	67-64-1	acetone	0,57	-1,10	8E+00	0	5
Soufrés	151-01-9	ethyl ethoxydithioformate	0,57	-0,66	3E-02	-2	5
Aldéhydes	78-84-2	2-methylpropanal	0,58	-0,74	5E-03	-3	5
Aldéhydes	123-38-6	propanal	0,60	-1,04	4E-03	-3	5
Aldéhydes	50-00-0	formaldehyde	0,61	-1,60	5E-01	-1	5
Alcènes	74-85-1	ethylene	0,61	-1,46	1E+02	2	5
Soufrés	10152-76-8	Methyl allyl sulfide	0,63	-0,68	5E-04	-4	5
Chlorés	7647-01-0	hydrogen chloride	0,63	-1,83	7E+00	0	5
Esters	107-31-3	methylformate	0,64	-1,12	3E+02	2	5
Aldéhydes	75-07-0	acetaldehyde	0,65	-1,33	8E-02	-2	5
Esters	79-20-9	methylacetate	0,66	-0,78	1E+00	0	5
Acides	79-09-4	acide propanoïque	0,68	-0,83	1E-02	-2	5
Acides	64-18-6	acide formique	0,71	-1,44	1E+00	0	5
Esters	109-94-4	Formate d'éthyle	0,72	-0,75	8E+00	0	5
Acides	64-19-7	acide acétique	0,72	-1,13	1E-01	-1	5
Alcools	78-92-2	2-butanol	0,73	-0,75	4E-01	-1	5
Alcools	75-65-0	2-methyl-2-propanol	0,74	-0,75	1E+01	1	5
Alcools	71-36-3	1-butanol	0,76	-0,72	5E-02	-2	5
Soufrés	7783-06-4	Hydrogen sulfide	0,76	-1,83	7E-04	-4	5
Alcools	67-63-0	isopropanol	0,78	-1,03	5E+01	1	5
Alcools	78-83-1	2-methylpropanol	0,79	-0,71	3E-02	-2	5
Soufrés	513-53-1	2-butanethiol	0,79	-0,74	1E-04	-4	5
Soufrés	68-11-1	mercaptoacetic acid	0,80	-0,91	8E-04	-4	5
Alcools	67-56-1	methanol	0,80	-1,57	4E+01	1	5
Soufrés	75-66-1	tert-butanethiol	0,80	-0,75	1E-04	-4	5
Alcools	71-23-8	1-propanol	0,80	-1,01	1E+00	0	5
Soufrés	75-18-3	dimethylsulfide(DMS)	0,81	-1,17	4E-03	-3	5
Soufrés	75-33-2	isopropyl mercaptane	0,82	-1,01	2E-05	-5	5
Soufrés	109-79-5	butanethiol	0,83	-0,68	1E-04	-4	5

Soufrés	74-93-1	methanethiol	0,83	-1,49	2E-04	-4	5
Alcools	64-17-5	ethanol	0,85	-1,26	9E-02	-2	5
Soufrés	107-03-9	n-propyl mercaptane	0,85	-0,96	4E-05	-5	5
Soufrés	513-44-0	1-isobutanethiol	0,86	-0,68	2E-05	-5	5
Alcane	106-97-8	butane	0,87	-0,80	3E+03	3	5
Soufrés	624-92-0	dimethyldisulfide (DMDS)	0,88	-0,90	6E-03	-3	5
Soufrés	75-08-1	Ethanethiol	0,89	-1,22	3E-05	-5	5
Alcane	74-98-6	Propane	0,91	-1,07	3E+03	3	5
Aldéhydes	2548-87-0	(E)-2-octenal	0,17	0,67	1E-02	-2	6
Cétones	110-93-0	6-methyl-5-hepten-2-one	0,19	0,59	3E-01	-1	6
Alcènes	124-11-8	1-nonene	0,35	0,81	3E-03	-3	6
Alcools	3391-86-4	1-octen-3-ol	0,37	0,62	5E-02	-2	6
Esters	106-63-8	isobutyl-2-propanoate	0,42	0,73	5E-03	-3	6
Cétones	111-13-7	2-octanone	0,42	0,64	4E-01	-1	6
Aldéhydes	123-05-7	2-ethylhexanal	0,44	0,69	8E-02	-2	6
Aldéhydes	124-13-0	octanal	0,45	0,70	3E-03	-3	6
Esters	108-65-6	1-methoxy-2-propyl acetate	0,57	0,56	2E-02	-2	6
Acides	111-14-8	Acide heptanoïque	0,57	0,63	3E-02	-2	6
Esters	638-11-9	isopropyl butanoate	0,62	0,67	3E-02	-2	6
Alcools	123-96-6	2-octanol	0,62	0,66	5E-02	-2	6
Esters	617-50-5	isopropyl isobutyrate	0,63	0,70	2E-01	-1	6
Alcools	111-87-5	1-octanol	0,64	0,69	2E-02	-2	6
Esters	105-66-8	propyl butyrate	0,65	0,71	6E-02	-2	6
Esters	539-82-2	pentanoate d'ethyl	0,65	0,72	1E-02	-2	6
Alcools	104-76-7	2-ethylhexanol	0,65	0,69	4E-01	-1	6
Esters	590-01-2	butyl propanoate	0,65	0,73	2E-01	-1	6
Esters	644-49-5	propyl isobutyrate	0,66	0,73	1E-02	-2	6
Ethers	5131-66-8	1-butoxy-2-propanol	0,67	0,60	1E+00	0	6
Esters	108-64-5	ethyl-3-methylbutanoate	0,68	0,72	7E-05	-5	6
Alcools	1653-40-3	6-Methyl-1-heptanol	0,68	0,70	5E-01	-1	6
Alcools	26952-21-6	isooctanol	0,68	0,70	5E-02	-2	6
Esters	540-42-1	isobutylpropionate	0,68	0,74	1E-01	-1	6
Esters	628-63-7	pentyl acetate (=amyl acetate)	0,68	0,72	1E-01	-1	6
Esters	123-92-2	3-methylbutyl acetate (isoamylacetate)	0,71	0,73	9E-02	-2	6
Esters	111-15-9	2-ethoxyethyl acetate	0,73	0,68	3E-01	-1	6
Alcane	111-84-2	nonane	0,73	0,85	1E+01	1	6
Alcane	111-65-9	octane	0,76	0,45	8E+00	0	6
Alcane	589-53-7	4-methylheptane	0,77	0,47	8E+00	0	6
Alcane	589-81-1	3-methylheptane	0,78	0,47	7E+00	0	6
Alcane	592-27-8	2-methyl-Heptane	0,79	0,47	5E-01	-1	6
Alcane	3522-94-9	2, 2, 5-trimethylhexane	0,81	0,93	5E+00	0	6
Alcane	540-84-1	isooctane	0,83	0,53	3E+00	0	6
Alcools	19700-21-1	Geosmine	-0,50	1,60	5E-05	-5	7
Aldéhydes	5392-40-5	3,7-dimethyl-2,6-octadienal	-0,12	1,54	8E-02	-2	7
Alcools	78-70-6	3,7-dimethylocta-1,6-dien-3-ol	0,07	1,46	4E-02	-2	7
Esters	32210-23-4	4-tert-butylcyclohexylacetate	0,09	2,50	9E-01	-1	7
Aldéhydes	3913-81-3	(E)-2-decenal	0,11	1,63	5E-03	-3	7
Alcools	1490-04-6	menthol (89-78-1)	0,12	1,06	7E+01	1	7
Alcools	89-78-1	menthol	0,12	1,06	7E+01	1	7
Aldéhydes	18829-56-6	(E)-non-2-enal	0,14	1,13	9E-03	-3	7
Aldéhydes	2463-53-8	2-nonenal	0,14	1,13	9E-03	-3	7
Aldéhydes	6728-35-4	6-nonenal	0,16	1,08	2E-02	-2	7
Cétones	112-12-9	2-undecanone	0,34	2,11	2E+00	0	7
Cétones	693-54-9	2-decanone	0,36	1,59	1E-01	-1	7
Aldéhydes	112-31-2	decanal	0,39	1,64	3E-03	-3	7
Alcools	106-22-9	A-citronellol	0,39	1,53	3E-01	-1	7
Cétones	821-55-6	nonan-2-one	0,39	1,10	2E-01	-1	7
Aldéhydes	124-19-6	nonanal	0,42	1,15	3E-03	-3	7
Cétones	108-83-8	2,6-dimethyl-4-heptanone	0,43	1,07	4E-01	-1	7
Acides	334-48-5	Acide decanoïque	0,48	2,10	5E-02	-2	7
Acides	112-05-0	Acide nonanoïque	0,52	1,57	2E-02	-2	7
Acides	124-07-2	Acide octanoïque	0,54	1,08	7E-03	-3	7
Esters	763-69-9	ethyl-3-ethoxy-propionate	0,56	1,02	1E-03	-3	7
Alcools	112-30-1	1-decanol	0,59	1,64	5E-03	-3	7
Esters	106-30-9	ethyl heptanoate	0,60	1,64	2E-01	-1	7
Alcools	143-08-8	nonan-1-ol	0,61	1,16	5E-03	-3	7
Esters	141-06-0	propyl valerate	0,62	1,15	2E-02	-2	7
Esters	109-21-7	butylbutyrate	0,62	1,14	3E-02	-2	7
Esters	109-19-3	butyl isovalerate	0,63	1,62	8E-02	-2	7
Esters	624-54-4	pentyl propanoate	0,63	1,16	9E-02	-2	7
Esters	106-27-4	isopentyl butanoate	0,63	1,62	7E-02	-2	7
Esters	138-22-7	butyl 2-hydroxypropanoate	0,63	1,02	3E-08	-8	7
Esters	97-87-0	butyl isobutyrate	0,64	1,16	1E-01	-1	7
Alcane	112-40-3	dodecane	0,65	2,30	3E+00	0	7
Alcools	106-21-8	3,7-dimethyloctanol	0,65	1,59	5E-03	-3	7
Esters	557-00-6	propyl isovalerate	0,65	1,15	3E-04	-4	7
Esters	539-90-2	isobutyl butyrate	0,65	1,15	9E-03	-3	7
Esters	142-92-7	Hexyl acetate	0,66	1,17	2E-01	-1	7
Esters	589-59-3	isobutyl isovalerate	0,66	1,62	3E-02	-2	7
Esters	97-85-8	isobutyl isobutyrate	0,67	1,17	4E-01	-1	7
Alcane	1120-21-4	undecane	0,68	1,78	7E+00	0	7
Alcane	124-18-5	decane	0,70	1,29	4E+00	0	7
Terpènes	7785-26-4	(1S,5S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-ene	-0,69	0,63	3E+00	0	8
Terpènes	7785-70-8	(1R,5R)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-ene	-0,69	0,63	2E+00	0	8

Terpènes	80-56-8	alpha-pinene	-0,69	0,63	2E-02	-2	8
Terpènes	127-91-3	Beta-pinene	-0,69	0,64	1E+00	0	8
Cétones	76-22-2	camphre	-0,62	0,89	2E-02	-2	8
Cétones	1195-79-5	Fenchone	-0,62	0,90	2E+00	0	8
Cétones	6485-40-1	(R)-5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-0,58	0,93	9E-02	-2	8
Cétones	99-49-0	5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-0,58	0,93	5E-01	-1	8
Ethers	470-82-6	eucalyptol	-0,50	0,79	1E-03	-3	8
Terpènes	99-85-4	gamma-terpinene	-0,45	0,71	5E-01	-1	8
Terpènes	138-86-3	Limonene	-0,42	0,73	6E-01	-1	8
Terpènes	5989-27-5	R-Limonene	-0,42	0,73	8E-02	-2	8
Cétones	89-82-7	(5R)-5-methyl-2-propan-2-ylidenecyclohexan-1-one	-0,37	0,97	8E-02	-2	8
Cétones	78-59-1	3,3,5-trimethyl-2-cyclohex-1-ene	-0,32	0,63	2E-03	-3	8
Aldéhydes	5910-87-2	(E, E)-2,4-nonadienal	-0,15	1,09	4E-04	-4	8
Alcools	98-55-5	alpha-terpineol	-0,13	0,96	2E-01	-1	8
Aldéhydes	30361-28-5	(E, E)-2,4-octadienal	-0,11	0,63	1E-02	-2	8
Acides	143-07-7	acide n-dodecanoïque	0,43	3,27	1E-01	-1	8
Esters	627-90-7	ethyl undecanoate	0,49	4,03	3E-04	-4	8
Esters	110-38-3	ethyl decanoate	0,52	3,36	1E-03	-3	8
Esters	112-17-4	decyl acetate	0,55	3,37	5E+01	1	8
Esters	25265-77-4	texanol	0,56	3,68	6E-01	-1	8

Annexe 17 DECOMPOSITION DES CLASSES SELON LES GROUPEMENTS FONCTIONNELS (350 MOLECULES)



Annexe 18 MODELISATION GLOBALE DU LOGARITHME DECIMAL DU SEUIL DE PERCEPTION DES 350 MOLECULES

Famille	CAS	Molécules	Log10(Y)	Log10(Ypred)	Division du jeu de	Erreur
Esters	138-22-7	butyl 2-hydroxypropanoate	-7,54	-1,62	train	5,91
Alcools	19700-21-	Geosmine	-4,31	-0,65	train	3,66
Azotés	75-50-3	triméthylamine	-3,59	-0,22	test	3,36
Cétones	127-41-3	4-(2,6,6-triméthyl-2-cyclohexen-1-yl)-3-buten-2-one	-4,26	-0,96	train	3,30
Alcane	74-98-6	Propane	3,43	0,39	train	3,04
Alcane	106-97-8	butane	3,46	0,43	test	3,03
Cétones	431-03-8	2,3-butanedione (diacetyl)	-3,93	-0,97	train	2,96
Soufrés	110-66-7	1-pentanethiol	-5,47	-2,72	train	2,74
Esters	108-64-5	ethyl-3-méthylbutanoate	-4,16	-1,43	train	2,73
Chlorés	75-09-2	dichloromethane	2,81	0,09	train	2,71
Soufrés	541-31-1	Isoamyl mercaptane	-5,48	-2,82	test	2,66
Ethers	470-82-6	eucalyptol	-2,84	-0,19	train	2,65
Soufrés	75-08-1	Ethanethiol	-4,49	-1,88	train	2,61
Alcènes	592-41-6	1-hexene	-3,28	-0,76	train	2,52
Esters	97-62-1	ethyl isobutyrate	-3,96	-1,50	test	2,46
Aromatiques	523-80-8	apiole	2,79	0,35	train	2,44
Esters	112-17-4	decyl acetate	1,68	-0,73	train	2,41
Aldéhydes	1205-17-0	7-méthyl-1,3-benzodioxole-5-propanal	-2,96	-0,55	test	2,41
Soufrés	75-33-2	isopropyl mercaptane	-4,72	-2,36	train	2,36
Soufrés	74-93-1	methanethiol	-3,74	-1,39	test	2,34
Aromatiques	106-44-5	4-méthylphenol (p.cresol)	-3,62	-1,33	train	2,29
Esters	627-90-7	ethyl undecanoate	-3,52	-1,24	train	2,29
Alcools	1490-04-6	menthol (89-78-1)	1,83	-0,43	train	2,27
Alcools	89-78-1	menthol	1,83	-0,43	train	2,27
Esters	557-00-6	propyl isovalerate	-3,48	-1,22	test	2,26
Cétones	78-59-1	3,3,5-triméthyl-2-cyclohex-1-ene	-2,77	-0,53	test	2,24
Esters	110-38-3	ethyl decanoate	-2,92	-0,72	train	2,20
Soufrés	7783-06-4	Hydrogen sulfide	-3,16	-1,09	train	2,07
Alcools	67-63-0	isopropanol	1,71	-0,35	test	2,06
Azotés	74-89-5	méthylamine	-2,11	-0,07	test	2,05
Aldéhydes	121-33-5	vanillin	-3,75	-1,71	test	2,04
Soufrés	107-03-9	n-propyl mercaptane	-4,40	-2,38	test	2,01
Aldéhydes	5910-87-2	(E, E)-2,4-nonadienal	-3,40	-1,39	train	2,01
Soufrés	111-31-9	1-hexanethiol	-4,14	-2,16	train	1,98
Aldéhydes	6728-26-3	2-hexenal (E)	-0,20	-2,17	test	1,97
Aromatiques	108-39-4	3-méthylphenol	-3,24	-1,32	train	1,93
Alcools	616-25-1	1-pentene-3-ol	0,63	-1,25	train	1,88
Alcools	75-65-0	2-méthyl-2-propanol	1,08	-0,78	train	1,86
Alcane	590-35-2	2,2-diméthylpentane	2,19	0,35	test	1,84
Aldéhydes	123-73-9	(E)-2-butenal	0,23	-1,60	train	1,83
Cétones	76-22-2	camphre	-1,80	0,03	test	1,83
Alcènes	124-11-8	1-monene	-2,55	-0,75	train	1,81
Esters	105-54-4	ethylbutyrate	-3,24	-1,46	train	1,78
Soufrés	513-44-0	1-isobutanethiol	-4,60	-2,83	train	1,78
Terpènes	80-56-8	alpha-pinene	-1,72	0,06	train	1,78
Esters	107-31-3	methylformate	2,51	0,75	train	1,77
Azotés	78-81-9	isobutylamine	-2,35	-0,65	train	1,70
Azotés	62-53-3	aniline	0,21	-1,48	train	1,69
Alcane	75-83-2	2,2-diméthylbutane	1,85	0,18	test	1,66
Aromatiques	95-48-7	2-méthylphenol	-2,92	-1,32	test	1,60
Soufrés	151-01-9	ethyl ethoxydithioformate	-1,52	-3,10	train	1,58
Alcools	75-85-4	2-méthyl-2-butanol	0,43	-1,13	train	1,56
Esters	140-88-5	ethyl acrylate	-3,00	-1,48	train	1,52
Alcools	143-08-8	nonan-1-ol	-2,28	-0,79	train	1,48
Aromatiques	105-05-5	1,4-diethylbenzene	-2,68	-1,21	test	1,47
Alcools	106-21-8	3,7-diméthyl-octanol	-2,34	-0,88	train	1,46
Alcools	67-56-1	methanol	1,59	0,15	test	1,44
Esters	763-69-9	ethyl-3-ethoxy-propionate	-2,96	-1,52	train	1,44
Alcools	112-30-1	1-decanol	-2,30	-0,88	train	1,42
Esters	108-21-4	isopropyl acetate	0,28	-1,12	train	1,40
Aldéhydes	590-86-3	3-méthylbutanal	-3,66	-2,26	train	1,39
Alcènes	111-66-0	1-octene	-2,14	-0,76	train	1,39
Azotés	124-40-3	diméthylamine	-1,51	-0,13	train	1,38
Esters	542-55-2	isobutyl formate	0,32	-1,05	test	1,37
Soufrés	75-66-1	tert-butanethiol	-3,96	-2,61	train	1,35
Alcènes	74-85-1	ethylene	2,10	0,76	train	1,33
Soufrés	2365-48-2	methyl 2-mercaptoacétate	-3,00	-1,67	test	1,33
Esters	80-62-6	methacrylate de methyl	-0,14	-1,47	train	1,33
Alcènes	115-11-7	2-méthyl-1-propene	1,36	0,03	train	1,33
Alcools	110-80-5	2-Ethoxyethanol	0,32	-0,99	train	1,31
Alcane	96-14-0	3-méthylpentane	1,49	0,19	train	1,31
Aldéhydes	18829-55-	(E)-2-heptenal	-0,73	-2,00	train	1,28
Acides	107-92-6	acide butanoïque (acide butyrique)	-3,31	-2,06	train	1,26
Ethers	110-00-9	furane	1,45	0,25	train	1,20

Soufrés	513-53-1	2-butanethiol	-3,96	-2,77	train	1,19
Chlorés	7647-01-0	hydrogen chloride	0,85	-0,34	test	1,18
Soufrés	110-02-1	tiophene	-2,22	-1,09	train	1,14
Esters	96-33-3	methyl acrylate	-1,96	-0,83	train	1,13
Ethers	5131-66-8	1-butoxy-2-propanol	0,11	-1,00	train	1,12
Azotés	94-68-8	N-ethyl-2-methylaniline	-0,85	-1,97	train	1,12
Azotés	83-34-1	skatole	-3,44	-2,33	train	1,11
Alcynes	74-86-2	acetylene	2,38	1,28	train	1,10
Cétones	112-12-9	2-undecanone	0,23	-0,86	test	1,09
Acides	503-74-2	acide 3-methylbutanoïque	-3,56	-2,47	train	1,09
Alcools	6032-29-7	2-pentanol	0,00	-1,09	test	1,09
Alcènes	111-67-1	2-octene	0,30	-0,75	train	1,06
Aromatiques	93-04-9	2-methoxynaphthalene	-3,00	-1,95	train	1,05
Esters	109-60-4	n-propyl acetate	0,00	-1,05	train	1,05
Alcènes	110-82-7	cyclohexane	1,57	0,53	train	1,04
Cétones	110-12-3	5-methyl-2-hexanone	-2,00	-0,96	train	1,04
Cétones	563-80-4	3-methyl-2-butanone	0,26	-0,78	train	1,04
Alcènes	565-59-3	2,3-dimethyl pentane	1,26	0,23	train	1,03
Aldéhydes	123-38-6	propanal	-2,44	-1,42	train	1,02
Azotés	5724-81-2	1-pyrroline	-1,10	-0,08	test	1,01
Alcènes	1120-21-4	undecane	0,98	-0,03	train	1,01
Alcools	64-17-5	ethanol	-1,03	-0,02	train	1,01
Aldéhydes	123-72-8	butanal	-2,91	-1,92	train	0,99
Chlorés	503-38-8	trichloromethyl chloroformate	-2,10	-1,10	train	0,99
Esters	540-88-5	acetate de 1,1-diméthylethyle	-0,47	-1,46	train	0,99
Cétones	107-87-9	2-pentanone	0,28	-0,70	test	0,98
Alcools	111-87-5	1-octanol	-1,73	-0,77	train	0,96
Alcènes	112-40-3	dodecane	1,07	0,11	train	0,96
Aldéhydes	124-13-0	octanal	-2,58	-1,62	test	0,96
Esters	539-82-2	pentanoate d'ethyl	-0,48	-1,43	test	0,95
Esters	141-78-6	acetate d'ethyle	0,46	-0,48	train	0,94
Azotés	75-31-0	isopropylamine	-1,22	-0,31	train	0,91
Alcools	1565-80-6	(S)-2-methyl-1-Butanol	-0,22	-1,13	train	0,90
Esters	106-36-5	n-propyl propionate	-0,55	-1,46	test	0,90
Azotés	75-04-7	ethylamine	-1,08	-0,18	test	0,90
Aromatiques	95-47-6	o-xylene	0,17	-0,72	train	0,90
Esters	547-63-7	methyl isobutyrate	-2,10	-1,20	train	0,89
Aldéhydes	124-19-6	nonanal	-2,59	-1,70	train	0,88
Esters	111-15-9	2-ethoxyethyl acetate	-0,59	-1,46	train	0,88
Alcènes	115-07-1	1-propene	1,34	0,48	train	0,87
Aromatiques	71-43-2	benzene	0,94	0,08	train	0,86
Cétones	108-10-1	4-methyl-2-pentanone	-0,14	-1,00	train	0,85
Esters	110-74-7	n-propyl formate	0,54	-0,31	train	0,85
Aromatiques	122-79-2	phenyl acetate	-0,70	-1,54	train	0,84
Esters	539-90-2	isobutyl butyrate	-2,03	-1,22	test	0,81
Esters	628-63-7	pentyl acetate (=amyl acetate)	-0,57	-1,36	train	0,79
Aromatiques	100-66-3	methoxybenzene	-0,25	-1,04	train	0,79
Esters	97-85-8	isobutyl isobutyrate	-0,36	-1,14	train	0,78
Cétones	67-64-1	acetone	0,89	0,11	train	0,78
Aromatiques	141-93-5	m-diethylbenzene	-0,41	-1,18	train	0,77
Soufrés	556-61-6	methyl isothiocyanate	0,71	-0,06	train	0,77
Aldéhydes	112-31-2	decanal	-2,55	-1,79	train	0,76
Alcènes	110-54-3	hexane	1,09	0,33	train	0,76
Alcènes	111-84-2	nonane	1,08	0,33	train	0,75
Aromatiques	95-63-6	1,2,4-trimethylbenzene	-0,19	-0,94	test	0,75
Cétones	591-78-6	2-hexanone	-0,22	-0,96	train	0,74
Alcènes	108-87-2	methylcyclohexane	-0,22	0,51	train	0,74
Soufrés	75-15-0	carbonedisulfide	-1,00	-0,26	test	0,73
Soufrés	2179-57-9	Diallyl disulfide	-2,89	-2,16	train	0,73
Alcènes	592-27-8	2-methyl-Heptane	-0,28	0,45	train	0,73
Azotés	91-22-5	quinoline	-1,52	-2,25	train	0,73
Esters	617-50-5	isopropyl isobutyrate	-0,72	-1,44	train	0,72
Cétones	78-93-3	2-butanone	0,41	-0,29	train	0,69
Aromatiques	90-05-1	2-methoxyphenol	-2,43	-1,74	test	0,69
Esters	590-01-2	butyl propanoate	-0,72	-1,41	train	0,69
Ethers	1634-04-4	methyl tert-butyl ether	-1,60	-0,93	train	0,68
Alcools	78-83-1	2-methylpropanol	-1,46	-0,80	train	0,66
Aromatiques	103-65-1	propylbenzene	-1,72	-1,07	train	0,65
Soufrés	68-11-1	mercaptoacetic acid	-3,10	-2,45	train	0,65
Cétones	110-93-0	6-methyl-5-hepten-2-one	-0,52	-1,16	train	0,64
Chlorés	75-36-5	ethanoyl chloride	-0,96	-0,33	train	0,63
Esters	106-63-8	isobutyl-2-propanoate	-2,33	-1,71	train	0,62
Azotés	120-72-9	indole	-2,85	-2,24	test	0,62
Chlorés	67-66-3	chloroforme	1,46	0,86	train	0,61
Terpènes	5989-27-5	R-Limonene	-1,08	-0,47	train	0,61
Ethers	109-99-9	tetrahydrofurane	0,35	-0,25	train	0,61
Esters	109-94-4	Formate d'ethyle	0,91	0,30	train	0,61
Alcools	106-22-9	A-citronellol	-0,55	-1,15	test	0,60
Alcools	71-41-0	1-pentanol	-0,45	-1,04	test	0,59
Alcools	26952-21-	isooctanol	-1,31	-0,73	train	0,58

Esters	142-92-7	Hexyl acetate	-1,81	-1,23	train	0,58
Esters	91-64-5	Coumarin	-2,00	-1,43	train	0,57
Esters	624-24-8	methyl valerate	-2,00	-1,43	train	0,57
Acides	109-52-4	acide pentanoïque	-2,96	-2,40	train	0,56
Alcools	123-96-6	2-octanol	-1,30	-0,75	test	0,55
Cétones	99-49-0	5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-0,30	-0,85	train	0,55
Aromatiques	611-14-3	2-ethyltoluene	-0,44	-0,99	train	0,55
Alcane	111-65-9	octane	0,90	0,36	train	0,54
Esters	644-49-5	propyl isobutyrate	-1,96	-1,42	train	0,54
Esters	592-84-7	n-butyl formate	-0,43	-0,96	train	0,53
Alcools	71-36-3	1-butanol	-1,28	-0,76	train	0,53
Aldéhydes	3913-81-3	(E)-2-decenal	-2,31	-1,80	train	0,51
Aldéhydes	123-11-5	4-methoxybenzaldehyde	-1,47	-1,97	test	0,50
Esters	105-37-3	ethyl propionate	-1,66	-1,16	train	0,50
Ethers	1708-29-8	2-5-dihydrofuran	-0,57	-0,08	train	0,49
Acides	79-09-4	acide propanoïque	-1,95	-1,46	train	0,49
Aldéhydes	96-17-3	2-methylbutanal	-2,78	-2,29	train	0,49
Esters	589-59-3	isobutyl isovalerate	-1,47	-0,98	test	0,49
Esters	556-24-1	methyl 3-methylbutanoate	-1,96	-1,47	test	0,49
Chlorés	108-90-7	chlorobenzene	0,00	-0,48	test	0,48
Cétones	111-13-7	2-octanone	-0,37	-0,85	test	0,48
Souffrés	592-88-1	Allyl sulfide	-3,00	-2,52	test	0,48
Acides	111-14-8	Acide heptanoïque	-1,48	-1,96	train	0,48
Azotés	107-10-8	propylamine	-0,82	-0,35	train	0,47
Esters	141-06-0	propyl valerate	-1,72	-1,25	train	0,47
Alcools	71-23-8	1-propanol	0,10	-0,36	train	0,46
Terpènes	7785-26-4	(1S, 5S)-2, 6,6-Trimethylbicyclo [3.1.1] hept-2-ene	0,52	0,06	train	0,46
Aromatiques	108-88-3	toluene	0,09	-0,38	train	0,46
Alcane	589-53-7	4-methylheptane	0,90	0,45	train	0,46
Alcane	124-18-5	decane	0,56	0,10	train	0,45
Alcools	111-76-2	2-butoxyethanol	-0,68	-1,13	train	0,45
Esters	625-55-8	isopropyl formate	0,00	-0,45	train	0,45
Aromatiques	108-95-2	phenol	-1,34	-0,89	test	0,45
Aldéhydes	123-05-7	2-ethylhexanal	-1,11	-1,55	train	0,43
Cétones	108-94-1	cyclohexanone	0,05	-0,39	train	0,43
Esters	540-42-1	isobutylpropionate	-0,96	-1,39	train	0,43
Esters	105-46-4	sec. butyl acetate	-1,92	-1,49	train	0,43
Azotés	134-20-3	methyl 2-aminobenzoate	-2,26	-1,83	train	0,43
Souffrés	10152-76-	Methyl allyl sulfide	-3,30	-2,87	train	0,43
Esters	106-30-9	ethyl heptanoate	-0,68	-1,10	train	0,42
Alcools	1653-40-3	6-Methyl-1-heptanol	-0,31	-0,73	test	0,42
Alcools	111-27-3	hexanol	-1,51	-1,10	train	0,42
Aldéhydes	30361-28-	(E, E)-2,4-octadienal	-1,92	-1,51	train	0,41
Esters	140-11-4	acetate de benzyle	-0,76	-1,17	train	0,41
Alcane	589-81-1	3-methylheptane	0,85	0,45	test	0,41
Alcools	10415-87-	3-methyl-1-phenyl-3-pentanol	-0,66	-1,06	test	0,40
Azotés	108-48-5	2,6-dimethylpyridine	-2,52	-2,12	train	0,40
Aromatiques	622-96-8	p-ethyltoluene	-1,39	-0,99	test	0,40
Aldéhydes	4313-03-5	(E, E)-2,4-heptadienal	-1,24	-1,64	train	0,40
Alcools	78-92-2	2-butanol	-0,38	-0,78	test	0,39
Cétones	89-82-7	(5R)-5-methyl-2-propan-2-ylidenecyclohexan-1-one	-1,08	-0,68	train	0,39
Aromatiques	108-38-3	m-xylene	-0,31	-0,70	test	0,39
Azotés	109-06-8	2-picoline	-2,00	-1,62	train	0,38
Alcane	78-78-4	2-methylbutane	0,58	0,20	train	0,38
Esters	623-42-7	methylbutyrate	-1,48	-1,10	train	0,38
Ethers	3777-69-3	2-pentylfuran	-0,57	-0,94	train	0,37
Aromatiques	488-23-3	1, 2, 3,4-tetramethylbenzene	-1,21	-0,85	test	0,36
Terpènes	78-79-5	isoprene	-0,89	-0,53	train	0,35
Aldéhydes	120-57-0	3,4-(Methylenedioxy) benzaldehyde	-2,13	-1,79	train	0,34
Acides	646-07-1	acide 4-methylpentanoïque	-2,72	-2,38	train	0,34
Aldéhydes	78-84-2	2-methylpropanal	-2,32	-1,98	train	0,34
Cétones	1195-79-5	Fenchone	0,36	0,02	train	0,34
Alcane	108-08-7	2,4-dimethyl pentane	0,59	0,25	train	0,34
Aldéhydes	66-25-1	hexanal	-2,49	-2,15	train	0,34
Azotés	109-73-9	butylamine	-0,29	-0,63	train	0,33
Azotés	13952-84-	2-butylamine	-0,29	-0,62	train	0,33
Cétones	108-83-8	2,6-dimethyl-4-heptanone	-0,40	-0,72	train	0,32
Esters	123-92-2	3-methylbutyl acetate (isoamylacetate)	-1,03	-1,35	test	0,32
Esters	97-87-0	butyl isobutyrate	-0,89	-1,21	test	0,32
Alcools	78-70-6	3,7-dimethylocta-1,6-dien-3-ol	-1,36	-1,06	test	0,30
Aromatiques	97-53-0	eugenol	-0,70	-1,00	train	0,30
Aromatiques	104-51-8	butylbenzene	-1,59	-1,29	train	0,30
Cétones	821-55-6	nonan-2-one	-0,61	-0,91	train	0,30
Esters	119-36-8	methyl salicylate	-1,00	-1,30	train	0,30
Aromatiques	2189-60-8	octylbenzene	-0,16	-0,45	test	0,29
Azotés	75-64-9	tert-butylamine	-0,29	-0,58	train	0,29
Alcane	110-83-8	cyclohexene	-0,63	-0,34	train	0,29
Esters	109-21-7	butylbutyrate	-1,55	-1,27	train	0,29
Azotés	123-75-1	pyrrolidine	0,00	-0,29	train	0,29
Azotés	126-98-7	methacrylonitrile	0,91	0,65	train	0,28

Acides	143-07-7	acide n-dodecanoïque	-1,00	-0,72	test	0,28
Alcools	104-76-7	2-ethylhexanol	-0,42	-0,70	train	0,28
Ethers	60-29-7	diethylether	-1,00	-0,73	train	0,27
Azotés	109-89-7	Ethanamine, N-ethyl-	-0,95	-0,67	test	0,27
Aldéhydes	75-07-0	acetaldehyde	-1,10	-0,83	test	0,27
Terpènes	7785-70-8	(1R, 5R)-2, 6, 6-Trimethylbicyclo [3.1.1] hept-2-ene	0,32	0,06	test	0,26
Aromatiques	106-42-3	p-xylene	-0,44	-0,71	train	0,26
Alcanes	109-66-0	pentane	0,61	0,36	train	0,26
Acides	124-07-2	Acide octanoïque	-1,96	-1,71	train	0,25
Terpènes	138-86-3	Limonene	-0,22	-0,47	train	0,25
Ethers	3299-32-9	2, 4, 5-trimethyl-1,3-dioxolane	-0,41	-0,65	train	0,24
Alcanes	589-34-4	3-methylhexane	0,53	0,29	test	0,24
Acides	64-19-7	acide acetique	-1,00	-0,76	test	0,24
Azotés	110-89-4	piperidine	-0,30	-0,53	train	0,23
Aromatiques	95-13-6	indene	-1,70	-1,47	test	0,23
Cétones	6485-40-1	(R)-5-Isopropenyl-2-methyl-2-cyclohexen-1-one	-1,07	-0,85	train	0,22
Alcanes	96-37-7	methylcyclopentane	0,76	0,55	train	0,22
Esters	105-66-8	propyl butyrate	-1,24	-1,45	train	0,21
Aromatiques	91-57-6	2-methyl naphtalene	-2,40	-2,19	train	0,21
Azotés	110-86-1	pyridine	-1,12	-0,91	train	0,21
Alcènes	106-99-0	1,3-butadiene	0,32	0,12	test	0,20
Aromatiques	60-12-8	2-phenylethanol	-1,82	-1,62	train	0,20
Acides	64-18-6	acide formique	0,18	-0,02	train	0,20
Azotés	121-69-7	N, N-dimethyl-benzenamine	-1,92	-2,11	train	0,19
Azotés	75-05-8	acetonitrile	1,34	1,16	test	0,19
Aldéhydes	110-62-3	pentanal	-2,01	-2,19	train	0,18
Esters	637-78-5	isopropyl propionate	-1,70	-1,52	train	0,18
Esters	624-54-4	pentyl propanoate	-1,07	-1,25	train	0,18
Aldéhydes	18829-56-	(E)-non-2-enal	-1,60	-1,78	test	0,17
Aldéhydes	2463-53-8	2-nonenal	-1,60	-1,78	train	0,17
Soufrés	110-01-0	tetrahydrothiophene	-2,66	-2,49	test	0,16
Aldéhydes	111-71-7	heptanal	-1,70	-1,85	test	0,16
Aromatiques	108-67-8	1, 3, 5-trimethylbenzene	-0,72	-0,87	train	0,15
Aldéhydes	5392-40-5	3, 7-dimethyl-2, 6-octadienal	-1,00	-1,15	train	0,15
Chlorés	79-01-6	trichloroethene	0,59	0,74	test	0,15
Cétones	110-43-0	2-heptanone	-1,11	-0,96	train	0,15
Ethers	271-89-6	2,3-benzofuran	-1,32	-1,46	train	0,14
Esters	108-65-6	1-methoxy-2-propyl acetate	-1,80	-1,66	train	0,14
Soufrés	109-79-5	butanethiol	-2,82	-2,69	train	0,14
Cétones	98-86-2	acetophenone	-1,89	-1,76	train	0,13
Azotés	104-90-5	5-Ethyl-2-methylpyridine	-2,40	-2,27	train	0,13
Aromatiques	91-20-3	naphtalene	-2,15	-2,03	train	0,12
Azotés	591-22-0	3,5-dimethylpyridine	-2,30	-2,18	train	0,12
Soufrés	352-93-2	Diethyl sulfide	-3,11	-2,99	train	0,12
Aromatiques	135-01-3	o-diethylbenzene	-1,28	-1,18	train	0,11
Acides	79-31-2	acide 2-methylpropanoïque	-2,27	-2,16	train	0,10
Azotés	121-44-8	triethylamine	-1,40	-1,29	train	0,10
Alcènes	106-98-9	1-butene	-0,08	0,02	train	0,10
Alcanes	617-78-7	3-ethylpentane	0,18	0,28	train	0,10
Aldéhydes	2548-87-0	(E)-2-octenal	-1,89	-1,79	train	0,10
Esters	106-27-4	isopentyl butanoate	-1,15	-1,05	train	0,10
Soufrés	75-18-3	dimethylsulfide(DMS)	-2,40	-2,30	train	0,10
Azotés	583-58-4	3,4-dimethylpyridine	-2,30	-2,20	train	0,10
Terpènes	5989-08-2	(1R, 2S, 7R, 8R)-2, 6, 6,9-tetramethyltricyclo [5.4.0.0 (2.8)] undec-9-ene	-0,04	-0,14	train	0,10
Esters	32210-23-	4-tert-butylcyclohexylacetate	-0,07	-0,17	train	0,10
Azotés	111-26-2	hexylamine	-1,21	-1,11	train	0,10
Soufrés	463-58-1	Carbonyl sulfide	-0,85	-0,75	train	0,10
Cétones	75-97-8	3,3-dimethyl-2-butanone	-0,74	-0,84	train	0,10
Aldéhydes	107-02-8	2-propenal (acrolein)	-1,16	-1,06	train	0,10
Acides	112-05-0	Acide nonanoïque	-1,70	-1,60	train	0,10
Esters	25265-77-	texanol	-0,24	-0,34	train	0,10
Aromatiques	98-82-8	1-methylethylbenzene	-1,13	-1,03	train	0,10
Aromatiques	119-64-2	1, 2, 3,4-tetrahydronaphtalene	-1,30	-1,20	train	0,10
Alcools	98-55-5	alpha-terpinol	-0,74	-0,64	train	0,10
Terpènes	99-85-4	gamma-terpinene	-0,33	-0,43	train	0,10
Alcools	111-70-6	1-heptanol	-1,01	-0,91	train	0,10
Aromatiques	94-59-7	saffrole	-1,00	-1,10	train	0,10
Aromatiques	1077-16-3	hexylbenzene	-1,54	-1,44	train	0,10
Terpènes	127-91-3	Beta-pinene	0,12	0,02	train	0,10
Aldéhydes	6728-35-4	6-nonenal	-1,64	-1,74	train	0,10
Soufrés	1679-09-0	2-methyl-2-butanethiol	-2,82	-2,72	train	0,10
Alcanes	3522-94-9	2, 2, 5-trimethylhexane	0,68	0,58	train	0,10
Azotés	7664-41-7	ammonia	0,10	0,00	train	0,10
Azotés	107-13-1	acrylonitrile	1,28	1,18	train	0,10
Soufrés	623-51-8	ethyl-2-mercaptoacetate	-2,10	-2,20	train	0,10
Acides	334-48-5	Acide decanoïque	-1,30	-1,20	train	0,10
Aromatiques	100-42-5	styrene	-1,23	-1,13	train	0,10
Iodés	75-47-8	triiodomethane	-1,22	-1,12	train	0,10
Aromatiques	104-55-2	3-phenyl-2-propenal	-1,62	-1,72	train	0,10

Chlorés	127-18-4	tetrachloroethylene	1,05	0,95	train	0,10
Cétones	693-54-9	2-decanone	-0,96	-1,06	train	0,10
Chlorés	98-88-4	benzoyl chloride	-1,92	-1,82	train	0,10
Bromés	75-25-2	tribromomethane	0,30	0,20	train	0,10
Soufrés	624-92-0	dimethyldisulfide (DMS)	-2,24	-2,34	train	0,10
Chlorés	56-23-5	carbon tetrachloride	1,46	1,36	train	0,10
Soufrés	7446-09-5	Sulfure dioxide	0,36	0,26	train	0,10
Aromatiques	104-46-1	anethole	-1,24	-1,34	train	0,10
Esters	93-89-0	ethyl benzoate	-1,30	-1,20	train	0,10
Cétones	120-92-3	cyclopentanone	-0,04	-0,14	train	0,10
Cétones	600-14-6	2,3-pentadione	-1,20	-1,30	train	0,10
Aromatiques	89-83-8	2-isopropenyl-5-methylphenol	-1,42	-1,32	train	0,10
Chlorés	71-55-6	1, 1,1-trichloroethane	0,72	0,64	train	0,09
Acides	142-62-1	acide hexanoïque	-2,24	-2,32	test	0,08
Alcanes	591-76-4	2-methylhexane	0,23	0,31	train	0,08
Chlorés	75-44-5	carbonyl chloride	-0,30	-0,38	train	0,08
Aromatiques	620-14-4	1-ethyl-3-methylbenzene	-1,06	-0,98	train	0,07
Esters	123-86-4	acetate de butyle	-1,29	-1,36	train	0,07
Alcènes	109-67-1	1-pentene	-0,54	-0,47	train	0,06
Esters	109-19-3	butyl isovalerate	-1,11	-1,05	train	0,06
Alcanes	292-64-8	cyclooctane	0,56	0,50	test	0,06
Alcanes	142-82-5	heptane	0,43	0,38	train	0,05
Esters	554-12-1	methyl propionate	-0,46	-0,51	train	0,05
Cétones	589-38-8	3-hexanone	-0,96	-1,01	train	0,05
Alcools	123-51-3	3-methyl-1-butanol	-1,05	-1,09	test	0,05
Aromatiques	100-41-4	ethylbenzene	-0,85	-0,81	train	0,05
Cétones	565-61-7	3-methyl-2-pentanone	-1,01	-1,04	test	0,04
Soufrés	110-81-6	Diethyl disulfide	-2,75	-2,71	train	0,03
Esters	79-20-9	methylacetate	0,13	0,16	train	0,03
Esters	110-19-0	acetate d'isobutyl	-1,42	-1,39	train	0,03
Alcanes	79-29-8	2, 3-dimethylbutane	0,18	0,20	train	0,02
Alcools	3391-86-4	1-octen-3-ol	-1,32	-1,30	train	0,02
Alcanes	540-84-1	isooctane	0,49	0,51	train	0,01
Aldéhydes	78-85-3	2-methyl-2-propenal	-1,60	-1,62	train	0,01
Alcènes	592-76-7	1-heptene	-0,85	-0,83	train	0,01
Aldéhydes	50-00-0	formaldehyde	-0,26	-0,26	train	0,01
Esters	638-11-9	isopropyl butanoate	-1,48	-1,49	train	0,00

Annexe 19 MOLECULES SUPPORTS PARMI LES 277 MOLECULES DU JEU DE CONSTRUCTION

butyl 2-hydroxypropanoate	propyl valerate	5-Isopropenyl-2-methyl-2-cyclohexen-1-one
1-pentanethiol	alpha-pinene	butylamine
isopropyl mercaptane	isopropyl propionate	2-butylamine
1-isobutanethiol	ethyl propionate	tert-butylamine
Ethanethiol	6-nonenal	2-methyl-Heptane
Geosmine	2-propenal (acrolein)	methoxybenzene
Propane	methyl tert-butyl ether	texanol
ethyl-3-methylbutanoate	butylbenzene	n-propyl acetate
1-hexanethiol	butylbutyrate	Limonene
2-butanethiol	hexylbenzene	methylcyclohexane
tert-butanethiol	quinoline	(S)-2-methyl-1-Butanol
2,3-butanedione (diacetyl)	ethyl ethoxydithioformate	2-hexanone
butanethiol	dimethylamine	4-methyl-2-pentanone
4-methylphenol (p.cresol)	hexanol	methacrylate de methyl
3-methylbutanal	Acide heptanoïque	4-tert-butylcyclohexylacetate
acide 3-methylbutanoïque	methylbutyrate	(1R,2S,7R,8R)-2,6,6,9-tetramethyltricyclo[5.4.0.0(2.8)]undec-9-ene
ethyl undecanoate	2-methylpropanol	pyrrolidine
skatole	2-isopropenyl-5-methylphenol	isopropyl formate
(E, E)-2,4-nonadienal	triethylamine	1-butoxy-2-propanol
acide butanoïque (acide butyrique)	methyl salicylate	cyclohexanone
Methyl allyl sulfide	2,3-benzofuran	isopropyl acetate
3-methylphenol	isooctanol	toluene
1-hexene	o-diethylbenzene	1-propanol
ethylbutyrate	1-butanol	ammonia
Hydrogen sulfide	1-methylethylbenzene	Beta-pinene
Diethyl sulfide	(E, E)-2,4-heptadienal	o-xylene
mercaptoacetic acid	anethole	acide formique
ethyl acrylate	isopropylamine	3-ethylpentane
2-methoxynaphtalene	triiodomethane	aniline
ethyl-3-ethoxy-propionate	hexylamine	3-methyl-2-butanone
acide pentanoïque	2,3-pentadione	2-octene
ethyl decanoate	isopentyl butanoate	tribromomethane
butanal	pyridine	2-Ethoxyethanol
Diallyl disulfide	2-ethylhexanal	tetrahydrofurane
eucalyptol	2-heptanone	Fenchone
2-methyl-2-butanethiol	3,7-dimethyl-2,6-octadienal	Sulfure dioxide
2-methylbutanal	R-Limonene	2-methyl-2-butanol
acide 4-methylpentanoïque	dichloromethane	2-butanone
nonanal	apiole	acetate d'ethyle
1-nonene	ethanol	dodecane
decanal	1-heptanol	(1S,5S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-ene
2,6-dimethylpyridine	diethylether	n-propyl formate
hexanal	2-decanone	decane
tiophene	isobutylpropionate	2-methylbutane
propanal	ethanoyl chloride	2,4-dimethyl pentane
2-methyl naphthalene	pentyl acetate (=amyl acetate)	pentane
5-Ethyl-2-methylpyridine	isoprene	1-pentene-3-ol
isobutylamine	N-ethyl-2-methylaniline	methyl isothiocyanate
3,7-dimethyloctanol	Carbonyl sulfide	methylcyclopentane
2-methylpropanal	propylamine	undecane
isobutyl-2-propanoate	Hexyl acetate	acetone
(E)-2-decenal	acetate de benzyle	octane
1-decanol	3,3-dimethyl-2-butanone	4-methylheptane
3,5-dimethylpyridine	alpha-terpineol	Formate d'ethyle
acide 2-methylpropanoïque	(E)-2-heptenal	methacrylonitrile
nonan-1-ol	butyl propanoate	benzene
methyl 2-aminobenzoate	isopropyl isobutyrate	tetrachloroethylene
safrrole	phenyl acetate	nonane
3,4-(Methylenedioxy) benzaldehyde	eugenol	2-methyl-2-propanol
1-octene	ethyl heptanoate	hexane
Acide octanoïque	2-butoxyethanol	2,3-dimethyl pentane
methyl isobutyrate	nonan-2-one	2-methyl-1-propene
trichloromethyl chloroformate	cyclohexene	acrylonitrile
pentanal	2-ethoxyethyl acetate	1-propene
5-methyl-2-hexanone	2-5-dihydrofuran	furane
2-picoline	2-pentylfuran	carbon tetrachloride
Coumarin	6-methyl-5-hepten-2-one	chloroforme
methyl valerate	(E)-2-butenal	3-methylpentane
methyl acrylate	acetate de 1,1-dimethylethyle	cyclohexane
propyl isobutyrate	2-ethyltoluene	decyl acetate
acide propanoïque	p-xylene	menthol (89-78-1)
sec.butyl acetate	n-butyl formate	menthol
(E, E)-2,4-octadienal	2-ethylhexanol	ethylene
benzoyl chloride	m-diethylbenzene	acetylene
acetophenone	2,4,5-trimethyl-1,3-dioxolane	methylformate
2-phenylethanol	2,6-dimethyl-4-heptanone	4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one
1-methoxy-2-propyl acetate	isobutyl isobutyrate	(5R)-5-methyl-2-propan-2-ylidenecyclohexan-1-one
1-octanol	gamma-terpinene	(R)-5-Isopropenyl-2-methyl-2-cyclohexen-1-one
propylbenzene	piperidine	

**Annexe 20 RESULTATS TOTAUX DE LA MODELISATION GLOBALE ET AFFINEE CONSTRUITE SUR
350 MOLECULES**

Famille	CAS	Molécules	Division du jeu de données pour le modèle global	Division du jeu de données pour les modèles locaux	Catégorie selon l'erreur du modèle global	SP de la littérature après nettoyage ou mesuré par Olentica (mg/m ³)	SP prédit par le modèle global (mg/m ³)	SP affiné par modèles locaux (mg/m ³)
Ethers	1708-29-8	2-5-dihydrofuran	train	train	Cat1	2,70E-01	8,36E-01	4,28E-01
Aldéhydes	96-17-3	2-methylbutanal	train	train	Cat1	1,68E-03	5,16E-03	2,66E-03
Acides	111-14-8	Acide heptanoïque	train	train	Cat1	3,30E-02	1,10E-02	2,08E-02
Esters	141-06-0	propyl valerate	train	train	Cat1	1,90E-02	5,58E-02	3,01E-02
Alcools	71-23-8	1-propanol	train	train	Cat1	1,26E+00	4,32E-01	7,91E-01
Terpènes	7785-26-4	(1S, 5S)-2, 6,6-Trimethylbicyclo [3.1.1] hept-2-ene	train	train	Cat1	3,30E+00	1,14E+00	2,08E+00
Alcane	589-53-7	4-methylheptane	train	train	Cat1	8,00E+00	2,79E+00	5,04E+00
Alcane	124-18-5	decane	train	train	Cat1	3,60E+00	1,26E+00	5,28E+00
Alcools	111-76-2	2-butoxyethanol	train	train	Cat1	2,10E-01	7,38E-02	1,32E-01
Aldéhydes	123-05-7	2-ethylhexanal	train	train	Cat1	7,73E-02	2,84E-02	4,88E-02
Cétones	108-94-1	cyclohexanone	train	train	Cat1	1,11E+00	4,08E-01	6,99E-01
Esters	105-46-4	sec. butyl acetate	train	train	Cat1	1,20E-02	3,21E-02	1,90E-02
Azotés	134-20-3	methyl 2-aminobenzoate	train	train	Cat1	5,48E-03	1,47E-02	8,68E-03
Souffrés	10152-76-	Methyl allyl sulfide	train	train	Cat1	5,00E-04	1,33E-03	7,93E-04
Alcools	111-27-3	hexanol	train	train	Cat1	3,08E-02	8,02E-02	4,88E-02
Aldéhydes	30361-28-	(E, E)-2,4-octadienal	train	train	Cat1	1,20E-02	3,11E-02	1,90E-02
Azotés	108-48-5	2,6-dimethylpyridine	train	train	Cat1	3,00E-03	7,60E-03	4,76E-03
Aldéhydes	4313-03-5	(E, E)-2,4-heptadienal	train	train	Cat1	5,70E-02	2,27E-02	3,60E-02
Cétones	89-82-7	(5R)-5-methyl-2-propan-2-ylidenecyclohexan-1-one	train	train	Cat1	8,40E-02	2,07E-01	1,33E-01
Azotés	109-06-8	2-picoline	train	train	Cat1	1,00E-02	2,40E-02	1,59E-02
Alcane	78-78-4	2-methylbutane	train	train	Cat1	3,80E+00	1,60E+00	2,40E+00
Esters	623-42-7	methylbutyrate	train	train	Cat1	3,32E-02	7,89E-02	5,26E-02
Terpènes	78-79-5	isoprene	train	train	Cat1	1,30E-01	2,93E-01	2,06E-01
Aldéhydes	120-57-0	3,4-(Methylenedioxy) benzaldehyde	train	train	Cat1	7,40E-03	1,63E-02	1,17E-02
Aldéhydes	78-84-2	2-methylpropanal	train	train	Cat1	4,78E-03	1,05E-02	4,28E-03
Cétones	1195-79-5	Fenchone	train	train	Cat1	2,30E+00	1,05E+00	1,45E+00
Alcane	108-08-7	2,4-dimethyl pentane	train	train	Cat1	3,90E+00	1,79E+00	2,46E+00
Azotés	109-73-9	butylamine	train	train	Cat1	5,10E-01	2,37E-01	3,22E-01
Azotés	13952-84-	2-butylamine	train	train	Cat1	5,10E-01	2,41E-01	3,22E-01
Cétones	108-83-8	2,6-dimethyl-4-heptanone	train	train	Cat1	4,01E-01	1,91E-01	2,53E-01
Aromatiques	97-53-0	eugenol	train	train	Cat1	2,00E-01	1,00E-01	1,26E-01
Cétones	821-55-6	nonan-2-one	train	train	Cat1	2,45E-01	1,24E-01	1,92E-01
Esters	119-36-8	methyl salicylate	train	train	Cat1	9,98E-02	5,06E-02	6,30E-02
Alcènes	110-83-8	cyclohexene	train	train	Cat1	2,35E-01	4,56E-01	3,73E-01
Esters	109-21-7	butylbutyrate	train	train	Cat1	2,80E-02	5,42E-02	3,70E-02
Azotés	123-75-1	pyrrolidine	train	train	Cat1	1,00E+00	5,17E-01	6,30E-01
Alcools	104-76-7	2-ethylhexanol	train	train	Cat1	3,81E-01	2,02E-01	2,40E-01
Ethers	60-29-7	diethylether	train	train	Cat1	1,00E-01	1,88E-01	1,59E-01
Aromatiques	106-42-3	p-xylene	train	train	Cat1	3,61E-01	1,97E-01	2,28E-01
Acides	124-07-2	Acide octanoïque	train	train	Cat1	1,10E-02	1,97E-02	1,74E-02
Terpènes	138-86-3	Limonene	train	train	Cat1	6,01E-01	3,36E-01	4,15E-01
Ethers	3299-32-9	2, 4,5-trimethyl-1,3-dioxolane	train	train	Cat1	3,90E-01	2,26E-01	2,47E-01
Azotés	110-89-4	piperidine	train	train	Cat1	5,00E-01	2,95E-01	3,87E-01
Alcane	96-37-7	methylcyclopentane	train	train	Cat1	5,80E+00	3,52E+00	4,82E+00
Esters	105-66-8	propyl butyrate	train	train	Cat1	5,80E-02	3,58E-02	3,66E-02
Aromatiques	91-57-6	2-methyl naphthalene	train	train	Cat1	4,00E-03	6,46E-03	6,34E-03
Aromatiques	60-12-8	2-phenylethanol	train	train	Cat1	1,53E-02	2,41E-02	2,17E-02
Acides	64-18-6	acide formique	train	train	Cat1	1,50E+00	9,52E-01	9,43E-01
Aldéhydes	110-62-3	pentanal	train	train	Cat1	9,73E-03	6,40E-03	6,14E-03
Esters	637-78-5	isopropyl propionate	train	train	Cat1	2,00E-02	3,04E-02	2,61E-02
Aldéhydes	18829-56-	(E)-non-2-enal	train	train	Cat1	2,50E-02	1,67E-02	1,57E-02
Aromatiques	108-67-8	1, 3,5-trimethylbenzene	train	train	Cat1	1,90E-01	1,34E-01	1,44E-01
Aldéhydes	5392-40-5	3,7-dimethyl-2,6-octadienal	train	train	Cat1	9,99E-02	7,13E-02	6,30E-02
Ethers	271-89-6	2,3-benzofuran	train	train	Cat1	4,80E-02	3,45E-02	3,03E-02
Esters	108-65-6	1-methoxy-2-propyl acetate	train	train	Cat1	1,60E-02	2,20E-02	1,95E-02
Souffrés	109-79-5	butanethiol	train	train	Cat1	1,50E-03	2,05E-03	2,37E-03
Azotés	104-90-5	5-Ethyl-2-methylpyridine	train	train	Cat1	4,00E-03	5,34E-03	3,84E-03
Aromatiques	91-20-3	naphthalene	train	train	Cat1	7,00E-03	9,30E-03	9,39E-03
Azotés	591-22-0	3,5-dimethylpyridine	train	train	Cat1	5,00E-03	6,61E-03	4,82E-03
Aromatiques	135-01-3	o-diethylbenzene	train	train	Cat1	5,20E-02	6,67E-02	8,24E-02
Acides	79-31-2	acide 2-methylpropanoïque	train	train	Cat1	5,40E-03	6,85E-03	3,41E-03
Azotés	121-44-8	triethylamine	train	train	Cat1	4,01E-02	5,08E-02	6,36E-02
Alcane	617-78-7	3-ethylpentane	train	train	Cat1	1,50E+00	1,89E+00	2,38E+00
Aldéhydes	2548-87-0	(E)-2-octenal	train	train	Cat1	1,30E-02	1,63E-02	2,06E-02
Esters	106-27-4	isopentyl butanoate	train	train	Cat1	7,00E-02	8,82E-02	5,18E-02
Azotés	583-58-4	3,4-dimethylpyridine	train	train	Cat1	5,00E-03	6,30E-03	4,59E-03
Terpènes	5989-08-2	(1R, 2S, 7R, 8R)-2,6,6,9-tetramethyltricyclo [5.4.0.0 (2.8)] undec-9-ene	train	train	Cat1	9,20E-01	7,30E-01	8,45E-01
Esters	32210-23-	4-tert-butylcyclohexylacetate	train	train	Cat1	8,60E-01	6,83E-01	5,42E-01
Souffrés	463-58-1	Carbonyl sulfide	train	train	Cat1	1,40E-01	1,76E-01	1,76E-01
Cétones	75-97-8	3,3-dimethyl-2-butanone	train	train	Cat1	1,80E-01	1,43E-01	2,21E-01

Aldéhydes	107-02-8	2-propenal (acrolein)	train	train	CatI	6,89E-02	8,67E-02	1,01E-01
Esters	25265-77-	texanol	train	train	CatI	5,80E-01	4,61E-01	3,66E-01
Aromatiques	98-82-8	1-méthylethylbenzene	train	train	CatI	7,39E-02	9,31E-02	1,06E-01
Aromatiques	119-64-2	1, 2, 3, 4-tetrahydronaphtalene	train	train	CatI	5,00E-02	6,29E-02	5,54E-02
Terpènes	99-85-4	gamma-terpinene	train	train	CatI	4,69E-01	3,73E-01	5,27E-01
Alcools	111-70-6	1-heptanol	train	train	CatI	9,68E-02	1,22E-01	8,42E-02
Aromatiques	94-59-7	safrole	train	train	CatI	9,98E-02	7,93E-02	6,29E-02
Terpènes	127-91-3	Beta-pinene	train	train	CatI	1,32E+00	1,05E+00	1,82E+00
Aldéhydes	6728-35-4	6-nonenal	train	train	CatI	2,30E-02	1,83E-02	3,33E-02
Soufrés	1679-09-0	2-méthyl-2-butanethiol	train	train	CatI	1,50E-03	1,89E-03	2,32E-03
Azotés	7664-41-7	ammonia	train	train	CatI	1,26E+00	1,00E+00	1,01E+00
Azotés	107-13-1	acrylonitrile	train	train	CatI	1,90E+01	1,51E+01	1,64E+01
Soufrés	623-51-8	ethyl-2-mercaptoacetate	train	train	CatI	8,00E-03	6,36E-03	6,63E-03
Aromatiques	100-42-5	styrene	train	train	CatI	5,92E-02	7,45E-02	7,57E-02
Iodés	75-47-8	triiodomethane	train	train	CatI	6,00E-02	7,55E-02	7,52E-02
Aromatiques	104-55-2	3-phenyl-2-propenal	train	train	CatI	2,40E-02	1,91E-02	2,20E-02
Cétones	693-54-9	2-decanone	train	train	CatI	1,10E-01	8,74E-02	1,06E-01
Chlorés	98-88-4	benzoyl chloride	train	train	CatI	1,20E-02	1,51E-02	1,90E-02
Bromés	75-25-2	tribromomethane	train	train	CatI	2,00E+00	1,59E+00	1,58E+00
Chlorés	56-23-5	carbon tetrachloride	train	train	CatI	2,90E+01	2,31E+01	1,83E+01
Soufrés	7446-09-5	Sulfure dioxide	train	train	CatI	2,30E+00	1,83E+00	1,82E+00
Aromatiques	104-46-1	anethole	train	train	CatI	5,70E-02	4,53E-02	3,91E-02
Cétones	120-92-3	cyclopentanone	train	train	CatI	9,11E-01	7,24E-01	1,08E+00
Cétones	600-14-6	2,3-pentadione	train	train	CatI	6,30E-02	5,01E-02	8,08E-02
Aromatiques	89-83-8	2-isopropenyl-5-methylphenol	train	train	CatI	3,80E-02	4,78E-02	4,33E-02
Alcane	591-76-4	2-méthylhexane	train	train	CatI	1,70E+00	2,04E+00	2,69E+00
Chlorés	75-44-5	carbonyl chloride	train	train	CatI	5,00E-01	4,19E-01	4,27E-01
Esters	123-86-4	acetate de butyle	train	train	CatI	5,13E-02	4,34E-02	3,64E-02
Alcènes	109-67-1	1-pentene	train	train	CatI	2,90E-01	3,36E-01	3,75E-01
Esters	109-19-3	butyl isovalerate	train	train	CatI	7,80E-02	8,94E-02	4,92E-02
Alcane	142-82-5	heptane	train	train	CatI	2,70E+00	2,40E+00	3,28E+00
Esters	554-12-1	methyl propionate	train	train	CatI	3,50E-01	3,12E-01	2,66E-01
Cétones	589-38-8	3-hexanone	train	train	CatI	1,09E-01	9,79E-02	1,48E-01
Aromatiques	100-41-4	ethylbenzene	train	train	CatI	1,41E-01	1,57E-01	1,66E-01
Soufrés	110-81-6	Diethyl disulfide	train	train	CatI	1,78E-03	1,93E-03	1,13E-03
Esters	110-19-0	acetate d'isobutyl	train	train	CatI	3,80E-02	4,05E-02	2,97E-02
Alcane	79-29-8	2, 3-dimethylbutane	train	train	CatI	1,50E+00	1,58E+00	2,28E+00
Alcools	3391-86-4	1-octen-3-ol	train	train	CatI	4,80E-02	4,98E-02	5,83E-02
Aldéhydes	78-85-3	2-méthyl-2-propenal	train	train	CatI	2,50E-02	2,42E-02	2,66E-02
Alcènes	592-76-7	1-heptene	train	train	CatI	1,43E-01	1,46E-01	2,26E-01
Aldéhydes	50-00-0	formaldehyde	train	train	CatI	5,45E-01	5,53E-01	4,90E-01
Acides	79-09-4	acide propanoïque	train	test	CatI	1,13E-02	3,49E-02	1,22E-02
Azotés	107-10-8	propylamine	train	test	CatI	1,50E-01	4,46E-01	6,54E-01
Aromatiques	108-88-3	toluene	train	test	CatI	1,22E+00	4,22E-01	4,08E-01
Esters	625-55-8	isopropyl formate	train	test	CatI	1,00E+00	3,56E-01	2,67E-01
Esters	540-42-1	isobutylpropanoate	train	test	CatI	1,10E-01	4,05E-02	3,57E-02
Esters	106-30-9	ethyl heptanoate	train	test	CatI	2,10E-01	7,96E-02	1,60E-02
Esters	140-11-4	acetate de benzyle	train	test	CatI	1,73E-01	6,78E-02	7,34E-02
Éthers	3777-69-3	2-pentylfuran	train	test	CatI	2,70E-01	1,16E-01	2,18E-01
Acides	646-07-1	acide 4-méthylpentanoïque	train	test	CatI	1,90E-03	4,18E-03	4,08E-03
Aldéhydes	66-25-1	hexanal	train	test	CatI	3,25E-03	7,03E-03	9,73E-03
Aromatiques	104-51-8	butylbenzene	train	test	CatI	2,57E-02	5,13E-02	6,56E-02
Azotés	75-64-9	tert-butylamine	train	test	CatI	5,10E-01	2,61E-01	4,97E-01
Azotés	126-98-7	methacrylonitrile	train	test	CatI	8,10E+00	4,22E+00	4,61E+00
Alcane	109-66-0	pentane	train	test	CatI	4,10E+00	2,27E+00	2,62E+00
Cétones	6485-40-1	(R)-5-Isopropenyl-2-méthyl-2-cyclohexen-1-one	train	test	CatI	8,50E-02	1,42E-01	8,65E-02
Azotés	110-86-1	pyridine	train	test	CatI	7,59E-02	1,22E-01	7,62E-02
Azotés	121-69-7	N, N-diméthyl-benzenamine	train	test	CatI	1,20E-02	7,73E-03	5,14E-03
Esters	624-54-4	pentyl propanoate	train	test	CatI	8,50E-02	5,63E-02	3,33E-02
Cétones	110-43-0	2-heptanone	train	test	CatI	7,81E-02	1,09E-01	2,23E-01
Cétones	98-86-2	acetophenone	train	test	CatI	1,30E-02	1,76E-02	1,80E-02
Soufrés	352-93-2	Diethyl sulfide	train	test	CatI	7,80E-04	1,03E-03	8,98E-04
Alcènes	106-98-9	1-butene	train	test	CatI	8,30E-01	1,05E+00	1,34E+00
Soufrés	75-18-3	dimethylsulfide(DMS)	train	test	CatI	4,00E-03	5,04E-03	4,02E-03
Azotés	111-26-2	hexylamine	train	test	CatI	6,20E-02	7,81E-02	1,64E-01
Acides	112-05-0	Acide nonanoïque	train	test	CatI	2,00E-02	2,52E-02	7,00E-03
Alcools	98-55-5	alpha-terpineol	train	test	CatI	1,80E-01	2,27E-01	2,64E-01
Aromatiques	1077-16-3	hexylbenzene	train	test	CatI	2,90E-02	3,65E-02	3,42E-02
Alcane	3522-94-9	2, 2, 5-triméthylhexane	train	test	CatI	4,80E+00	3,81E+00	1,04E+01
Acides	334-48-5	Acide decanoïque	train	test	CatI	5,00E-02	6,29E-02	6,02E-03
Chlorés	127-18-4	tetrachloroethylene	train	test	CatI	1,13E+01	8,98E+00	7,33E+00
Soufrés	624-92-0	dimethyldisulfide (DMS)	train	test	CatI	5,69E-03	4,52E-03	4,09E-03
Esters	93-89-0	ethyl benzoate	train	test	CatI	5,00E-02	6,29E-02	7,05E-02
Chlorés	71-55-6	1, 1,1-trichloroethane	train	test	CatI	5,30E+00	4,35E+00	5,36E+00
Aromatiques	620-14-4	1-ethyl-3-méthylbenzene	train	test	CatI	8,80E-02	1,05E-01	1,15E-01
Esters	79-20-9	methylacetate	train	test	CatI	1,36E+00	1,46E+00	1,31E+00
Alcane	540-84-1	isooctane	train	test	CatI	3,10E+00	3,21E+00	5,87E+00
Esters	638-11-9	isopropyl butanoate	train	test	CatI	3,30E-02	3,27E-02	2,61E-02
Esters	589-59-3	isobutyl isovalerate	test	train	CatI	3,40E-02	1,04E-01	5,39E-02
Esters	556-24-1	methyl 3-méthylbutanoate	test	train	CatI	1,10E-02	3,36E-02	1,74E-02
Cétones	111-13-7	2-octanone	test	train	CatI	4,24E-01	1,40E-01	2,68E-01
Soufrés	592-88-1	Allyl sulfide	test	train	CatI	1,00E-03	3,01E-03	1,58E-03
Aromatiques	108-95-2	phenol	test	train	CatI	4,59E-02	1,29E-01	7,27E-02

Alcools	1653-40-3	6-Methyl-1-heptanol	test	train	Cat1	4,90E-01	1,87E-01	3,09E-01
Alcanes	589-81-1	3-methylheptane	test	train	Cat1	7,10E+00	2,79E+00	5,01E+00
Alcools	10415-87-1	3-methyl-1-phenyl-3-pentanol	test	train	Cat1	2,20E-01	8,67E-02	1,39E-01
Alcools	78-92-2	2-butanol	test	train	Cat1	4,15E-01	1,68E-01	2,62E-01
Aromatiques	488-23-3	1, 2, 3,4-tetramethylbenzene	test	train	Cat1	6,10E-02	1,41E-01	9,67E-02
Esters	97-87-0	butyl isobutyrate	test	train	Cat1	1,30E-01	6,21E-02	8,20E-02
Alcools	78-70-6	3,7-dimethylocta-1,6-dien-3-ol	test	train	Cat1	4,40E-02	8,79E-02	5,63E-02
Aromatiques	2189-60-8	octylbenzene	test	train	Cat1	6,90E-01	3,51E-01	1,09E+00
Acides	143-07-7	acide n-dodecanoïque	test	train	Cat1	1,00E-01	1,90E-01	6,31E-02
Aldéhydes	75-07-0	acetaldehyde	test	train	Cat1	7,88E-02	1,46E-01	1,25E-01
Terpènes	7785-70-8	(1R, 5R)-2, 6,6-Trimethylbicyclo [3.1.1] hept-2-ene	test	train	Cat1	2,10E+00	1,14E+00	2,08E+00
Acides	64-19-7	acide acetique	test	train	Cat1	1,01E-01	1,73E-01	6,36E-02
Aromatiques	95-13-6	indene	test	train	Cat1	2,00E-02	3,38E-02	3,17E-02
Alcènes	106-99-0	1,3-butadiene	test	train	Cat1	2,10E+00	1,33E+00	2,10E+00
Azotés	75-05-8	acetonitrile	test	train	Cat1	2,20E+01	1,44E+01	1,99E+01
Aldéhydes	2463-53-8	2-nonenal	test	train	Cat1	2,50E-02	1,67E-02	1,57E-02
Soufrés	110-01-0	tetrahydrothiophene	test	train	Cat1	2,20E-03	3,20E-03	3,48E-03
Chlorés	79-01-6	trichloroethene	test	train	Cat1	3,90E+00	5,45E+00	6,16E+00
Acides	142-62-1	acide hexanoïque	test	train	Cat1	5,82E-03	4,82E-03	9,23E-03
Cétones	565-61-7	3-methyl-2-pentanone	test	train	Cat1	9,80E-02	9,02E-02	1,04E-01
Chlorés	108-90-7	chlorobenzene	test	test	Cat1	1,00E+00	3,29E-01	2,83E-01
Aromatiques	622-96-8	p-ethyltoluene	test	test	Cat1	4,10E-02	1,03E-01	1,12E-01
Aromatiques	108-38-3	m-xylene	test	test	Cat1	4,86E-01	1,98E-01	2,26E-01
Esters	123-92-2	3-methylbutyl acetate (isoamylacetate)	test	test	Cat1	9,31E-02	4,44E-02	3,15E-02
Azotés	109-89-7	Ethanamine, N-ethyl-	test	test	Cat1	1,13E-01	2,12E-01	1,84E-01
Alcanes	589-34-4	3-methylhexane	test	test	Cat1	3,40E+00	1,97E+00	2,51E+00
Aldéhydes	111-71-7	heptanal	test	test	Cat1	2,01E-02	1,40E-02	2,24E-02
Alcanes	292-64-8	cyclooctane	test	test	Cat1	3,60E+00	3,17E+00	3,83E+00
Alcools	123-51-3	3-methyl-1-butanol	test	test	Cat1	8,99E-02	8,10E-02	8,06E-02
Aldéhydes	123-72-8	butanal	train	train	Cat2	1,23E-03	1,21E-02	4,90E-03
Esters	540-88-5	acetate de 1,1-dimethylethyle	train	train	Cat2	3,40E-01	3,50E-02	8,55E-02
Alcools	111-87-5	1-octanol	train	train	Cat2	1,86E-02	1,71E-01	7,40E-02
Esters	141-78-6	acetate d'ethyl	train	train	Cat2	2,91E+00	3,31E-01	7,31E-01
Alcools	1565-80-6	(S)-2-methyl-1-Butanol	train	train	Cat2	6,00E-01	7,47E-02	1,51E-01
Esters	547-63-7	methyl isobutyrate	train	train	Cat2	8,00E-03	6,24E-02	3,18E-02
Aldéhydes	124-19-6	nonanal	train	train	Cat2	2,59E-03	1,97E-02	1,03E-02
Esters	111-15-9	2-ethoxyethyl acetate	train	train	Cat2	2,60E-01	3,47E-02	6,53E-02
Aromatiques	71-43-2	benzene	train	train	Cat2	8,81E+00	1,20E+00	2,21E+00
Cétones	108-10-1	4-methyl-2-pentanone	train	train	Cat2	7,21E-01	1,01E-01	1,89E-01
Esters	110-74-7	n-propyl formate	train	train	Cat2	3,50E+00	4,91E-01	8,80E-01
Esters	628-63-7	pentyl acetate (=amyl acetate)	train	train	Cat2	2,70E-01	4,34E-02	6,78E-02
Aromatiques	100-66-3	methoxybenzene	train	train	Cat2	5,60E-01	9,15E-02	1,41E-01
Cétones	67-64-1	acetone	train	train	Cat2	7,75E+00	1,28E+00	1,95E+00
Aromatiques	141-93-5	m-diethylbenzene	train	train	Cat2	3,90E-01	6,56E-02	9,79E-02
Soufrés	556-61-6	methyl isothiocyanate	train	train	Cat2	5,10E+00	8,66E-01	1,28E+00
Alcanes	110-54-3	hexane	train	train	Cat2	1,24E+01	2,15E+00	3,11E+00
Alcanes	111-84-2	nonane	train	train	Cat2	1,20E+01	2,14E+00	3,01E+00
Alcanes	108-87-2	methylcyclohexane	train	train	Cat2	6,00E-01	3,27E+00	2,39E+00
Soufrés	2179-57-9	Diallyl disulfide	train	train	Cat2	1,30E-03	6,99E-03	5,18E-03
Azotés	91-22-5	quinoline	train	train	Cat2	3,00E-02	5,64E-03	7,54E-03
Esters	617-50-5	isopropyl isobutyrate	train	train	Cat2	1,90E-01	3,65E-02	4,77E-02
Cétones	78-93-3	2-butanone	train	train	Cat2	2,54E+00	5,15E-01	1,05E+00
Esters	590-01-2	butyl propanoate	train	train	Cat2	1,90E-01	3,87E-02	8,52E-02
Ethers	1634-04-4	methyl tert-butyl ether	train	train	Cat2	2,50E-02	1,18E-01	9,95E-02
Alcools	78-83-1	2-methylpropanol	train	train	Cat2	3,49E-02	1,59E-01	1,39E-01
Soufrés	68-11-1	mercaptoacetic acid	train	train	Cat2	8,00E-04	3,54E-03	3,18E-03
Cétones	110-93-0	6-methyl-5-hepten-2-one	train	train	Cat2	3,00E-01	6,95E-02	9,02E-02
Chlorés	75-36-5	ethanoyl chloride	train	train	Cat2	1,10E-01	4,71E-01	4,38E-01
Chlorés	67-66-3	chloroforme	train	train	Cat2	2,92E+01	7,19E+00	7,36E+00
Terpènes	5989-27-5	R-Limonene	train	train	Cat2	8,29E-02	3,36E-01	3,30E-01
Esters	109-94-4	Formate d'ethyl	train	train	Cat2	8,10E+00	2,01E+00	4,08E+00
Esters	142-92-7	Hexyl acetate	train	train	Cat2	1,55E-02	5,82E-02	6,16E-02
Esters	91-64-5	Coumarin	train	train	Cat2	1,00E-02	3,72E-02	3,98E-02
Esters	624-24-8	methyl valerate	train	train	Cat2	1,01E-02	3,71E-02	4,01E-02
Cétones	99-49-0	5-Isopropenyl-2-methyl-2-cyclohexen-1-one	train	train	Cat2	5,00E-01	1,42E-01	1,44E-01
Aromatiques	611-14-3	2-ethyltoluene	train	train	Cat2	3,60E-01	1,02E-01	1,27E-01
Esters	644-49-5	propyl isobutyrate	train	train	Cat2	1,10E-02	3,82E-02	4,16E-02
Esters	592-84-7	n-butyl formate	train	train	Cat2	3,70E-01	1,10E-01	1,82E-01
Alcools	71-36-3	1-butanol	train	train	Cat2	5,21E-02	1,76E-01	2,07E-01
Esters	105-37-3	ethyl propionate	train	train	Cat2	2,19E-02	6,97E-02	8,73E-02
Chlorés	503-38-8	trichloromethyl chloroformate	train	test	Cat2	8,00E-03	7,86E-02	8,12E-02
Alcanes	112-40-3	dodecane	train	test	Cat2	1,18E+01	1,30E+00	1,40E+00
Azotés	75-31-0	isopropylamine	train	test	Cat2	6,00E-02	4,92E-01	3,96E-01
Aromatiques	95-47-6	o-xylene	train	test	Cat2	1,49E+00	1,89E-01	2,38E-01
Alcènes	115-07-1	1-propene	train	test	Cat2	2,20E+01	3,00E+00	3,06E+00
Aromatiques	122-79-2	phenyl acetate	train	test	Cat2	2,00E-01	2,90E-02	4,33E-02
Esters	97-85-8	isobutyl isobutyrate	train	test	Cat2	4,40E-01	7,24E-02	4,24E-02
Aldéhydes	112-31-2	decanal	train	test	Cat2	2,81E-03	1,62E-02	1,00E-02
Cétones	591-78-6	2-hexanone	train	test	Cat2	6,03E-01	1,10E-01	2,72E-01
Alcanes	592-27-8	2-methyl-Heptane	train	test	Cat2	5,20E-01	2,79E+00	2,95E+00
Aromatiques	103-65-1	propylbenzene	train	test	Cat2	1,90E-02	8,47E-02	1,08E-01
Esters	106-63-8	isobutyl-2-propanoate	train	test	Cat2	4,70E-03	1,97E-02	2,05E-02

Ethers	109-99-9	tetrahydrofurane	train	test	Cat2	2,26E+00	5,60E-01	6,73E-01
Alcools	26952-21-	isoocétanol	train	test	Cat2	4,90E-02	1,87E-01	1,23E-01
Acides	109-52-4	acide pentanoïque	train	test	Cat2	1,10E-03	4,01E-03	4,06E-03
Alcane	111-65-9	octane	train	test	Cat2	8,00E+00	2,29E+00	3,19E+00
Aldéhydes	3913-81-3	(E)-2-decenal	train	test	Cat2	4,90E-03	1,60E-02	1,62E-02
Cétones	107-87-9	2-pentanone	test	train	Cat2	1,90E+00	1,99E-01	4,77E-01
Aldéhydes	124-13-0	octanal	test	train	Cat2	2,63E-03	2,38E-02	1,05E-02
Esters	539-82-2	pentanoate d'éthyl	test	train	Cat2	3,29E-01	3,72E-02	8,27E-02
Esters	106-36-5	n-propyl propionate	test	train	Cat2	2,80E-01	3,49E-02	7,02E-02
Azotés	75-04-7	éthylamine	test	train	Cat2	8,30E-02	6,60E-01	3,30E-01
Esters	539-90-2	isobutyl butyrate	test	train	Cat2	9,40E-03	6,08E-02	3,74E-02
Aromatiques	95-63-6	1,2,4-triméthylbenzène	test	train	Cat2	6,43E-01	1,15E-01	1,62E-01
Souffrés	75-15-0	carbonedisulfide	test	train	Cat2	1,01E-01	5,45E-01	4,02E-01
Azotés	120-72-9	indole	test	train	Cat2	1,40E-03	5,81E-03	5,57E-03
Alcools	106-22-9	A-citronellol	test	train	Cat2	2,80E-01	7,04E-02	7,04E-02
Alcools	71-41-0	1-pentanol	test	train	Cat2	3,55E-01	9,08E-02	8,91E-02
Alcools	123-96-6	2-octanol	test	train	Cat2	5,00E-02	1,76E-01	1,39E-01
Aldéhydes	123-11-5	4-méthoxybenzaldehyde	test	train	Cat2	3,40E-02	1,07E-02	1,15E-02
Aromatiques	90-05-1	2-méthoxyphénol	test	test	Cat2	3,70E-03	1,82E-02	2,64E-02
Alcools	143-08-8	nonan-1-ol	train	train	Cat3	5,30E-03	1,61E-01	6,67E-03
Alcools	106-21-8	3,7-diméthyl-1-octanol	train	train	Cat3	4,59E-03	1,32E-01	5,78E-03
Esters	763-69-9	éthyl-3-éthoxy-propionate	train	train	Cat3	1,10E-03	3,02E-02	1,38E-03
Esters	108-21-4	isopropyl acetate	train	train	Cat3	1,90E+00	7,53E-02	1,51E+00
Aldéhydes	590-86-3	3-méthylbutanal	train	train	Cat3	2,20E-04	5,46E-03	2,77E-04
Alcènes	111-66-0	1-octène	train	train	Cat3	7,18E-03	1,75E-01	9,04E-03
Souffrés	75-66-1	tert-butanolthiol	train	train	Cat3	1,10E-04	2,47E-03	1,38E-04
Alcènes	74-85-1	éthylène	train	train	Cat3	1,25E+02	5,80E+00	9,93E+01
Esters	80-62-6	méthacrylate de méthyl	train	train	Cat3	7,30E-01	3,40E-02	5,80E-01
Alcènes	115-11-7	2-méthyl-1-propène	train	train	Cat3	2,30E+01	1,07E+00	1,83E+01
Alcools	110-80-5	2-éthoxyéthanol	train	train	Cat3	2,10E+00	1,02E-01	1,67E+00
Aldéhydes	18829-55-	(E)-2-heptenal	train	train	Cat3	1,87E-01	9,93E-03	1,49E-01
Acides	107-92-6	acide butanoïque (acide butyrique)	train	train	Cat3	4,88E-04	8,78E-03	6,15E-04
Ethers	110-00-9	furane	train	train	Cat3	2,80E+01	1,77E+00	2,22E+01
Souffrés	110-02-1	thiophène	train	train	Cat3	5,99E-03	8,20E-02	7,54E-03
Esters	96-33-3	méthyl acrylate	train	train	Cat3	1,10E-02	1,48E-01	1,38E-02
Azotés	94-68-8	N-éthyl-2-méthylaniline	train	train	Cat3	1,40E-01	1,07E-02	1,11E-01
Azotés	83-34-1	skatole	train	train	Cat3	3,66E-04	4,73E-03	4,61E-04
Alcynes	74-86-2	acétylène	train	train	Cat3	2,40E+02	1,89E+01	1,91E+02
Acides	503-74-2	acide 3-méthylbutanoïque	train	train	Cat3	2,77E-04	3,41E-03	3,49E-04
Alcènes	111-67-1	2-octène	train	train	Cat3	2,00E+00	1,76E-01	1,59E+00
Esters	109-60-4	n-propyl acetate	train	train	Cat3	9,99E-01	8,95E-02	7,93E-01
Alcane	110-82-7	cyclohexane	train	train	Cat3	3,72E+01	3,38E+00	2,96E+01
Cétones	110-12-3	5-méthyl-2-héxanone	train	train	Cat3	9,90E-03	1,09E-01	1,25E-02
Alcane	565-59-3	2,3-diméthylpentane	train	train	Cat3	1,80E+01	1,70E+00	1,43E+01
Aldéhydes	123-38-6	propanal	train	train	Cat3	3,60E-03	3,76E-02	4,54E-03
Alcools	64-17-5	éthanol	train	train	Cat3	9,36E-02	9,49E-01	1,18E-01
Alcools	112-30-1	1-décaneol	train	test	Cat3	5,00E-03	1,33E-01	2,66E-02
Azotés	124-40-3	diméthylamine	train	test	Cat3	3,06E-02	7,37E-01	7,24E-01
Alcane	96-14-0	3-méthylpentane	train	test	Cat3	3,10E+01	1,53E+00	3,06E+00
Souffrés	513-53-1	2-butanethiol	train	test	Cat3	1,10E-04	1,70E-03	1,76E-03
Ethers	5131-66-8	1-butoxy-2-propanol	train	test	Cat3	1,30E+00	9,89E-02	9,50E-02
Aromatiques	93-04-9	2-méthoxynaphtalène	train	test	Cat3	1,00E-03	1,13E-02	1,12E-02
Cétones	563-80-4	3-méthyl-2-butanone	train	test	Cat3	1,80E+00	1,65E-01	1,37E-01
Alcane	1120-21-4	undécane	train	test	Cat3	9,59E+00	9,38E-01	9,27E-01
Alcools	67-56-1	méthanol	test	train	Cat3	3,91E+01	1,42E+00	3,10E+01
Esters	542-55-2	isobutyl formate	test	train	Cat3	2,10E+00	8,89E-02	1,67E+00
Chlorés	7647-01-0	hydrogène chlorure	test	train	Cat3	7,00E+00	4,62E-01	5,56E+00
Alcools	6032-29-7	2-pentanol	test	train	Cat3	1,00E+00	8,16E-02	7,94E-01
Azotés	5724-81-2	1-pyrroline	test	train	Cat3	8,00E-02	8,26E-01	1,01E-01
Aromatiques	105-05-5	1,4-diéthylbenzène	test	test	Cat3	2,10E-03	6,20E-02	5,74E-02
Souffrés	2365-48-2	méthyl 2-mercaptoacétate	test	test	Cat3	1,00E-03	2,15E-02	2,13E-02
Cétones	112-12-9	2-undécanone	test	test	Cat3	1,70E+00	1,37E-01	1,26E-01
Esters	138-22-7	butyl 2-hydroxypropanoate	train	train	Cat4	2,90E-08	2,38E-02	9,18E-08
Cétones	127-41-3	4-(2,6,6-triméthyl-2-cyclohexen-1-yl)-3-butène-2-one	train	train	Cat4	5,49E-05	1,09E-01	1,73E-04
Alcane	74-98-6	Propane	train	train	Cat4	2,70E+03	2,47E+00	8,54E+02
Cétones	431-03-8	2,3-butanedione (diacétyle)	train	train	Cat4	1,17E-04	1,07E-01	3,71E-04
Souffrés	110-66-7	1-pentanethiol	train	train	Cat4	3,40E-06	1,88E-03	1,08E-05
Esters	108-64-5	éthyl-3-méthylbutanoate	train	train	Cat4	6,90E-05	3,68E-02	2,18E-04
Ethers	470-82-6	eucalyptol	train	train	Cat4	1,44E-03	6,42E-01	4,56E-03
Souffrés	75-08-1	Ethanethiol	train	train	Cat4	3,24E-05	1,32E-02	1,03E-04
Aromatiques	523-80-8	apiolène	train	train	Cat4	6,10E+02	2,22E+00	1,93E+02
Esters	112-17-4	decyl acetate	train	train	Cat4	4,76E+01	1,85E-01	1,50E+01
Souffrés	75-33-2	isopropyl mercaptane	train	train	Cat4	1,90E-05	4,39E-03	6,01E-05
Aromatiques	106-44-5	4-méthylphénol (p-cresol)	train	train	Cat4	2,40E-04	4,70E-02	7,59E-04
Alcools	1490-04-6	menthol (89-78-1)	train	train	Cat4	6,80E+01	3,68E-01	2,15E+01
Alcools	89-78-1	menthol	train	train	Cat4	6,80E+01	3,68E-01	2,15E+01
Esters	110-38-3	éthyl decanoate	train	train	Cat4	1,20E-03	1,90E-01	3,80E-03
Souffrés	7783-06-4	Hydrogène sulfure	train	train	Cat4	6,97E-04	8,18E-02	2,20E-03
Aldéhydes	5910-87-2	(E)-2,4-nonadienal	train	train	Cat4	4,00E-04	4,06E-02	1,26E-03
Aromatiques	108-39-4	3-méthylphénol	train	train	Cat4	5,70E-04	4,84E-02	9,30E-04
Alcools	616-25-1	1-pentène-3-ol	train	train	Cat4	4,30E+00	5,64E-02	1,36E+00
Aldéhydes	123-73-9	(E)-2-butenal	train	train	Cat4	1,70E+00	2,51E-02	5,37E-01
Esters	105-54-4	éthylbutyrate	train	train	Cat4	5,78E-04	3,46E-02	1,40E-03

Soufrés	513-44-0	1-isobutanethiol	train	train	Cat4	2,50E-05	1,50E-03	7,91E-05
Alcènes	592-41-6	1-hexene	train	train	Cat4	5,22E-04	1,81E-01	7,00E-04
Azotés	78-81-9	isobutylamine	train	train	Cat4	4,50E-03	2,24E-01	1,42E-02
Azotés	62-53-3	aniline	train	train	Cat4	1,64E+00	3,32E-02	5,18E-01
Soufrés	151-01-9	ethyl ethoxydithioformate	train	train	Cat4	3,00E-02	7,94E-04	9,48E-03
Alcools	75-85-4	2-methyl-2-butanol	train	train	Cat4	2,68E+00	7,37E-02	8,46E-01
Esters	140-88-5	ethyl acrylate	train	train	Cat4	1,00E-03	3,29E-02	3,16E-03
Alcools	19700-21-	Geosmine	train	test	Cat4	4,90E-05	2,22E-01	2,03E-01
Chlorés	75-09-2	dichloromethane	train	test	Cat4	6,40E+02	1,24E+00	1,16E+00
Terpènes	80-56-8	alpha-pinene	train	test	Cat4	1,92E-02	8,79E-01	7,90E-01
Esters	627-90-7	ethyl undecanoate	train	test	Cat4	3,00E-04	5,80E-02	1,33E-04
Soufrés	111-31-9	1-hexanethiol	train	test	Cat4	7,20E-05	6,89E-03	3,69E-05
Alcools	75-65-0	2-methyl-2-propanol	train	test	Cat4	1,21E+01	1,65E-01	7,06E-01
Alcènes	124-11-8	1-nonene	train	test	Cat4	2,80E-03	1,79E-01	1,73E-01
Esters	107-31-3	methylformate	train	test	Cat4	3,25E+02	5,56E+00	4,71E+00
Azotés	75-50-3	trimethylamine	test	train	Cat4	2,60E-04	5,98E-01	8,21E-04
Soufrés	541-31-1	isoamyl mercaptane	test	train	Cat4	3,30E-06	1,52E-03	1,04E-05
Esters	97-62-1	ethyl isobutyrate	test	train	Cat4	1,10E-04	3,16E-02	3,48E-04
Soufrés	74-93-1	methanethiol	test	train	Cat4	1,83E-04	4,06E-02	5,80E-04
Esters	557-00-6	propyl isovalerate	test	train	Cat4	3,30E-04	6,07E-02	2,99E-04
Alcools	67-63-0	isopropanol	test	train	Cat4	5,12E+01	4,48E-01	1,62E+01
Aldéhydes	121-33-5	vanillin	test	train	Cat4	1,80E-04	1,96E-02	5,68E-04
Soufrés	107-03-9	n-propyl mercaptane	test	train	Cat4	4,00E-05	4,13E-03	1,26E-04
Aldéhydes	6728-26-3	2-hexenal (E)	test	train	Cat4	6,30E-01	6,76E-03	1,99E-01
Alcènes	590-35-2	2, 2-dimethylpentane	test	train	Cat4	1,56E+02	2,25E+00	4,93E+01
Cétones	76-22-2	camphre	test	train	Cat4	1,60E-02	1,08E+00	5,06E-02
Alcènes	75-83-2	2,2-dimethylbutane	test	train	Cat4	7,00E+01	1,53E+00	2,21E+01
Alcènes	106-97-8	butane	test	test	Cat4	2,88E+03	2,70E+00	1,47E+01
Aldéhydes	1205-17-0	7-methyl-1,3-benzodioxole-5-propanal	test	test	Cat4	1,10E-03	2,82E-01	2,80E-01
Cétones	78-59-1	3, 3,5-trimethyl-2-cyclohex-1-ene	test	test	Cat4	1,70E-03	2,98E-01	1,16E-01
Azotés	74-89-5	methylamine	test	test	Cat4	7,69E-03	8,60E-01	8,16E-01
Aromatiques	95-48-7	2-methylphenol	test	test	Cat4	1,20E-03	4,76E-02	6,54E-04

Annexe 21 DONNEES PERMETTANT D'EVALUER LA JUSTESSE ET LA PRECISION DE LA FABRICATION D'UN MELANGE LIQUIDE

N° CAS	Molécule	Famille	Volume (μL)	densité	Masse théorique (mg)	Masse expérimentale (mg)					Moyenne masse expérimentale (mg)	SD	RSD	Erreur relative	Erreur relative (%)
						Essai A	Essai B	Essai C	Essai D	Essai E					
109-66-0	pentane	Alcanes	100	0.649	64,9	45,5	43	39,8	36,8	41,5	41,3	3,3	7,941	0,36	36,3
60-29-7	diethylether	ethers	100	0.714	71,4	45,4	26,8	64,8	67,3	40,6	49,0	17,0	34,781	0,31	31,4
75-18-3	diméthylsulfide(DMS)	Soufrés	100	0.84	84,0	80,8	52,9	58,5	79	52,6	64,8	14,0	21,670	0,23	22,9
79-29-8	2, 3-diméthylbutane	Alcanes	100	0.662	66,2	57,2	56,4	50,6	56,7	54,9	55,2	2,7	4,876	0,17	16,7
75-83-2	2,2-diméthylbutane	Alcanes	100	0.675	67,5	61,3	58,8	60,7	52,4	50,4	56,7	5,0	8,804	0,16	16,0
71-36-3	1-butanol	Alcools	100	0.81	81,0	54,4	74,2	75,1	75,4	62,8	68,4	9,4	13,775	0,16	15,6
1634-04-4	méthyl tert-butyl ether	Ethers	100	0.75	75,0	61,1	67,3	63,8	64,7	62,8	63,9	2,3	3,605	0,15	14,7
109-99-9	tetrahydrofurane	Ethers	100	0.89	89,0	NA	67	72,7	82	84,8	76,6	8,2	10,756	0,14	13,9
96-14-0	3-méthylpentane	Alcanes	100	0.673	67,3	61,4	57,4	56,4	56,8	58,4	58,1	2,0	3,449	0,14	13,7
67-64-1	acetone	Cétones	100	0.79	79,0	69,4	71,1	72,1	63,7	70,2	69,3	3,3	4,745	0,12	12,3
110-54-3	hexane	Alcanes	100	0.659	65,9	60,4	55,3	64,1	59,4	57,9	59,4	3,2	5,463	0,10	9,8
78-93-3	2-butanone	Cétones	100	0.786	78,6	76,2	71,1	70,3	74	71,9	72,7	2,4	3,291	0,08	7,5
141-78-6	acetate d'éthyle	Esters	100	0.898	89,8	86,4	80,9	81,8	83,5	84	83,3	2,1	2,556	0,07	7,2
110-83-8	cyclohexene	Alcènes	100	0.823	82,3	77,9	77,3	75,1	76,4	77,6	76,9	1,1	1,474	0,07	6,6
592-76-7	1-heptene	Alcènes	100	0.697	69,7	66,6	64,4	65,4	67,1	65,4	65,8	1,1	1,632	0,06	5,6
111-66-0	1-octene	Alcènes	100	0.72	72,0	68,5	67,4	71,5	67,1	67,8	68,5	1,8	2,598	0,05	4,9
105-37-3	ethyl propionate	Esters	100	0.888	88,8	87,0	83,5	84,1	83,2	85,4	84,6	1,6	1,850	0,05	4,7
100-41-4	ethylbenzene	Aromatiques	100	0.868	86,8	83,9	83,2	83,2	80,5	82,9	82,7	1,3	1,577	0,05	4,7
591-78-6	2-hexanone	Cétones	100	0.81	81,0	77,6	76,9	78,7	77,1	77,9	77,6	0,7	0,918	0,04	4,1
108-10-1	4-méthyl-2-pentanone	Cétones	100	0.8	80,0	77,3	80,2	76	74,5	75,8	76,8	2,2	2,819	0,04	4,0
100-42-5	styrene	Aromatiques	100	0.902	90,2	86,7	85,3	86,9	87,1	87,2	86,6	0,8	0,893	0,04	3,9
106-42-3	p-xylene	Aromatiques	100	0.86	86,0	81,6	81,6	84,8	81,6	83,9	82,7	1,5	1,862	0,04	3,8
111-84-2	nonane	Alcanes	100	0.724	72,4	69,0	68,1	70,3	69,4	72,4	69,8	1,6	2,341	0,04	3,5
123-86-4	acetate de butyle	Esters	100	0.886	88,6	84,1	86,8	88,5	85,2	83,2	85,6	2,1	2,480	0,03	3,4
589-38-8	3-hexanone	Cétones	100	0.803	80,3	78,2	76,8	77,5	78,2	77,7	77,7	0,6	0,747	0,03	3,3
592-41-6	1-hexene	Alcènes	100	0.678	67,8	62,5	57,5	76,3	70,6	62,3	65,8	7,5	11,399	0,03	2,9
96-17-3	2-méthylbutanal	Aldéhydes	100	0.791	79,1	82,0	80,2	77,3	86,1	80	81,1	3,3	4,007	0,03	2,6
111-65-9	octane	Alcanes	100	0.703	70,3	67,8	66,2	69,7	66,1	72,9	68,5	2,8	4,148	0,03	2,5
108-38-3	m-xylene	Aromatiques	100	0.864	86,4	85,9	85	84,6	82,2	83,8	84,3	1,4	1,656	0,02	2,4
1120-21-4	undecane	Alcanes	300	0.743	222,9	214,3	214,9	220,9	215,8	225,2	218,2	4,7	2,151	0,02	2,1
110-43-0	2-heptanone	Cétones	100	0.808	80,8	78,1	77,6	82	79,4	78,5	79,1	1,7	2,199	0,02	2,1
124-18-5	decane	Alcanes	300	0.735	220,5	218,0	212,1	212,4	217,5	219,6	215,9	3,4	1,593	0,02	2,1
104-76-7	2-ethylhexanol	Alcools	300	0.821	246,3	227,0	268,6	228,8	268,6	263,8	251,4	21,5	8,559	0,02	2,1
95-63-6	1,2,4-triméthylbenzene	Aromatiques	300	0.88	264,0	254,7	258,9	258,3	262,2	260	258,8	2,7	1,059	0,02	2,0
108-88-3	toluene	Aromatiques	100	0.871	87,1	85,6	82,8	89,3	85,3	84,3	85,5	2,4	2,819	0,02	1,9
142-82-5	heptane	Alcanes	100	0.695	69,5	65,2	63,5	66,1	80,4	66	68,2	6,9	10,078	0,02	1,8
624-24-8	méthyl valerate	Esters	100	0.886	88,6	87,8	86,6	86,8	86	88,4	87,1	1,0	1,108	0,02	1,7
108-83-8	2,6-diméthyl-4-heptanone	Cétones	300	0.811	243,3	231,7	233	244,6	249,6	240,3	239,8	7,6	3,170	0,01	1,4
105-54-4	ethylbutyrate	Esters	100	0.886	88,6	86,1	89,4	92,4	83,4	86,4	87,5	3,4	3,940	0,01	1,2
112-31-2	decanal	Aldéhydes	300	0.83	249,0	243,1	241,1	255,2	266,7	250,1	251,2	10,3	4,103	0,01	0,9
95-47-6	o-xylene	Aromatiques	100	0.879	87,9	90,8	88,5	88,6	83,8	83,9	87,1	3,1	3,585	0,01	0,9
98-86-2	acetophenone	Cétones	300	1,0266	308,0	311,6	310,8	303,8	318,6	302,5	309,5	6,5	2,109	0,00	0,5
590-86-3	3-méthylbutanal	Aldéhydes	100	0.791	79,1	77,3	72,8	80,1	81,7	84,8	79,3	4,6	5,738	0,00	0,3
623-42-7	méthylbutyrate	Esters	100	0.891	89,1	92,4	83,3	88,1	93,8	88,4	89,2	4,1	4,624	0,00	0,1
123-05-7	2-ethylhexanal	Aldéhydes	300	0.809	242,7	247,7	237,2	238,6	243,2	247,7	242,9	4,9	2,029	0,00	0,1

Annexe 22 EXEMPLE D'APPLICATION DU PROTOCOLE DE VOLATILISATION D'UN MELANGE LIQUIDE SYNTHETIQUE EN MELANGE GAZEUX SYNTHETIQUE

Dans cette annexe est expliquée la manière dont un mélange gazeux est obtenu à partir d'un mélange liquide. Il est décrit particulièrement l'obtention des concentrations.

Soit N le nombre de molécules d'un mélange liquide synthétique. Un volume théorique de chacune des molécules (v_{ith}) est prélevé et additionné entre-elles de manière successive. Chaque volume individuel est pesé à son ajout dans le mélange (m_{iexp}). Puis, à l'aide de la masse volumique (μ_i), il est calculé le volume expérimental correspondant à cette pesée (v_{iexp}) selon l'Équation 15.

$$v_{iexp} = \frac{m_{iexp}}{\mu_i}$$

v_{iexp} : le volume expérimental de la molécule i dans le mélange (m^3)
 m_{iexp} : la masse expérimentale de la molécule i dans le mélange (kg)
 μ_i : la masse volumique de la molécule i (kg/m^3)

Équation 15 : Calcul du volume expérimental de chaque molécule i introduit dans le mélange liquide synthétique à partir de la pesée expérimentale

Le volume total expérimental du mélange est calculé (V_{exp}) selon l'Équation 16.

$$V_{exp} = \sum_1^N v_{iexp}$$

v_{iexp} : le volume de la molécule i (m^3)
 N : le nombre de molécules dans le mélange

Équation 16 : Définition du volume total expérimental du mélange liquide

La concentration massique d'une molécule i en solution (c_{mi}) est obtenue selon l'Équation 17.

$$c_{mi} = \frac{m_{iexp} \times 10^6}{V_{exp}}$$

c_{mi} : la concentration massique de la molécule i (mg/m^3)
 m_{iexp} : la masse de la molécule i (mg)
 V_{exp} : le volume total du mélange (m^3)

Équation 17 : Calcul de la concentration massique d'une molécule i dans le mélange liquide synthétique

Un volume V_i de ce mélange liquide est prélevé pour être volatilisé dans un sac Nalophan® de 40 litres (V_{sac}). Il est alors obtenu la concentration des molécules i dans le mélange gazeux (c_{mgi}) récupéré dans ce sac de 40 litres. Cette dernière concentration est obtenue selon l'Équation 18 qui repose sur la conservation de la quantité de matière.

$$c_{mgi} = \frac{c_{mi} \times V_i}{V_{sac}}$$

c_{mgi} : la concentration de la molécule i dans le mélange gazeux (mg/m^3)
 c_{mi} : la concentration de la molécule i dans le mélange liquide (mg/m^3)
 V_i : le volume du mélange liquide prélevé pour volatilisation (m^3)
 V_{sac} : le volume total du mélange liquide (m^3)

Équation 18 : Calcul de la concentration massique d'une molécule i dans le mélange gazeux synthétique

Le résultat obtenu est présenté pour le mélange 1 en Tableau 70. Celui-ci est obtenu par volatilisation d'un volume V_i de 5 μ L dans un sac de volume V_{sac} de 40L.

Tableau 70 : Concentrations des molécules du mélange 1 synthétique à l'état liquide et gazeux

N° CAS	Molécule	v_{ith} (m^3)	v_{iexp} (m^3)	μ_i (kg/m^3)	m_{iexp} (kg)	c_{mi} (mg/m^3)	c_{mgi} (mg/m^3)
79-29-8	2, 3-dimethylbutane	1.00E-07	1.31E-07	6.62E+02	8.66E-05	1.44E+07	1.80
111-84-2	nonane	1.00E-07	3.22E-07	7.24E+02	2.33E-04	3.88E+07	4.85
109-66-0	pentane	1.00E-07	1.20E-07	6.49E+02	7.76E-05	1.29E+07	1.62
1120-21-4	undecane	3.00E-07	1.12E-07	7.43E+02	8.32E-05	1.39E+07	1.73
96-14-0	3-methylpentane	1.00E-07	3.85E-07	6.73E+02	2.59E-04	4.31E+07	5.39
110-54-3	hexane	1.00E-07	1.03E-07	6.59E+02	6.81E-05	1.13E+07	1.42
111-65-9	octane	1.00E-07	4.42E-07	7.03E+02	3.11E-04	5.18E+07	6.47
124-18-5	decane	3.00E-07	1.20E-07	7.35E+02	8.85E-05	1.47E+07	1.84
75-83-2	2,2-dimethylbutane	1.00E-07	3.57E-07	6.75E+02	2.41E-04	4.02E+07	5.02
142-82-5	heptane	1.00E-07	3.09E-07	6.95E+02	2.15E-04	3.58E+07	4.47
592-41-6	1-hexene	1.00E-07	1.22E-07	6.78E+02	8.28E-05	1.38E+07	1.72
111-66-0	1-octene	1.00E-07	1.18E-07	7.20E+02	8.50E-05	1.42E+07	1.77
592-76-7	1-heptene	1.00E-07	1.25E-07	6.97E+02	8.68E-05	1.45E+07	1.81
110-83-8	cyclohexene	1.00E-07	9.33E-08	8.23E+02	7.68E-05	1.28E+07	1.60
71-36-3	1-butanol	1.00E-07	6.83E-08	8.10E+02	5.53E-05	9.21E+06	1.15
104-76-7	2-ethylhexanol	3.00E-07	9.04E-08	8.21E+02	7.42E-05	1.24E+07	1.54
590-86-3	3-methylbutanal	1.00E-07	5.44E-08	7.91E+02	4.30E-05	7.16E+06	0.90
96-17-3	2-methylbutanal	1.00E-07	8.51E-08	7.91E+02	6.73E-05	1.12E+07	1.40
112-31-2	decanal	3.00E-07	3.23E-08	8.30E+02	2.68E-05	4.46E+06	0.56
123-05-7	2-ethylhexanal	3.00E-07	6.54E-08	8.09E+02	5.29E-05	8.81E+06	1.10
100-42-5	styrene	1.00E-07	9.24E-08	9.02E+02	8.33E-05	1.39E+07	1.73
108-38-3	m-xylene	1.00E-07	9.66E-08	8.64E+02	8.35E-05	1.39E+07	1.74
95-63-6	1,2,4-trimethylbenzene	3.00E-07	8.27E-08	8.80E+02	7.28E-05	1.21E+07	1.52
100-41-4	ethylbenzene	1.00E-07	9.24E-08	8.68E+02	8.02E-05	1.34E+07	1.67
108-88-3	toluene	1.00E-07	1.03E-07	8.71E+02	8.94E-05	1.49E+07	1.86
95-47-6	o-xylene	1.00E-07	9.28E-08	8.79E+02	8.16E-05	1.36E+07	1.70
106-42-3	p-xylene	1.00E-07	7.70E-08	8.60E+02	6.62E-05	1.10E+07	1.38
98-86-2	acetophenone	3.00E-07	6.93E-08	1.03E+03	7.11E-05	1.18E+07	1.48
110-43-0	2-heptanone	1.00E-07	7.97E-08	8.08E+02	6.44E-05	1.07E+07	1.34
589-38-8	3-hexanone	1.00E-07	1.01E-07	8.03E+02	8.09E-05	1.35E+07	1.68
108-83-8	2,6-dimethyl-4-heptanone	3.00E-07	8.31E-08	8.11E+02	6.74E-05	1.12E+07	1.40
591-78-6	2-hexanone	1.00E-07	9.49E-08	8.10E+02	7.69E-05	1.28E+07	1.60
78-93-3	2-butanone	1.00E-07	3.02E-07	7.86E+02	2.37E-04	3.95E+07	4.94
108-10-1	4-methyl-2-pentanone	1.00E-07	8.38E-08	8.00E+02	6.70E-05	1.12E+07	1.39
67-64-1	acetone	1.00E-07	3.40E-07	7.90E+02	2.69E-04	4.47E+07	5.59
105-54-4	ethylbutyrate	1.00E-07	8.02E-08	8.86E+02	7.11E-05	1.18E+07	1.48
624-24-8	methyl valerate	1.00E-07	6.37E-08	8.86E+02	5.64E-05	9.39E+06	1.17
105-37-3	ethyl propionate	1.00E-07	9.03E-08	8.88E+02	8.02E-05	1.34E+07	1.67
623-42-7	methylbutyrate	1.00E-07	6.44E-08	8.91E+02	5.74E-05	9.56E+06	1.20
123-86-4	acetate de butyle	1.00E-07	6.49E-08	8.86E+02	5.75E-05	9.58E+06	1.20
141-78-6	acetate d'ethyle	1.00E-07	9.50E-08	8.98E+02	8.53E-05	1.42E+07	1.78
1634-04-4	methyl tert-butyl ether	1.00E-07	1.03E-07	7.50E+02	7.73E-05	1.29E+07	1.61
60-29-7	diethylether	1.00E-07	8.89E-08	7.14E+02	6.35E-05	1.06E+07	1.32
109-99-9	tetrahydrofurane	1.00E-07	2.38E-07	8.90E+02	2.12E-04	3.53E+07	4.42
75-18-3	dimethylsulfide(DMS)	1.00E-07	7.00E-08	8.40E+02	5.88E-05	9.79E+06	1.22

Il est alors obtenu un sac contenant un mélange gazeux de composition et de concentration en molécules connues prêt à être analysé.

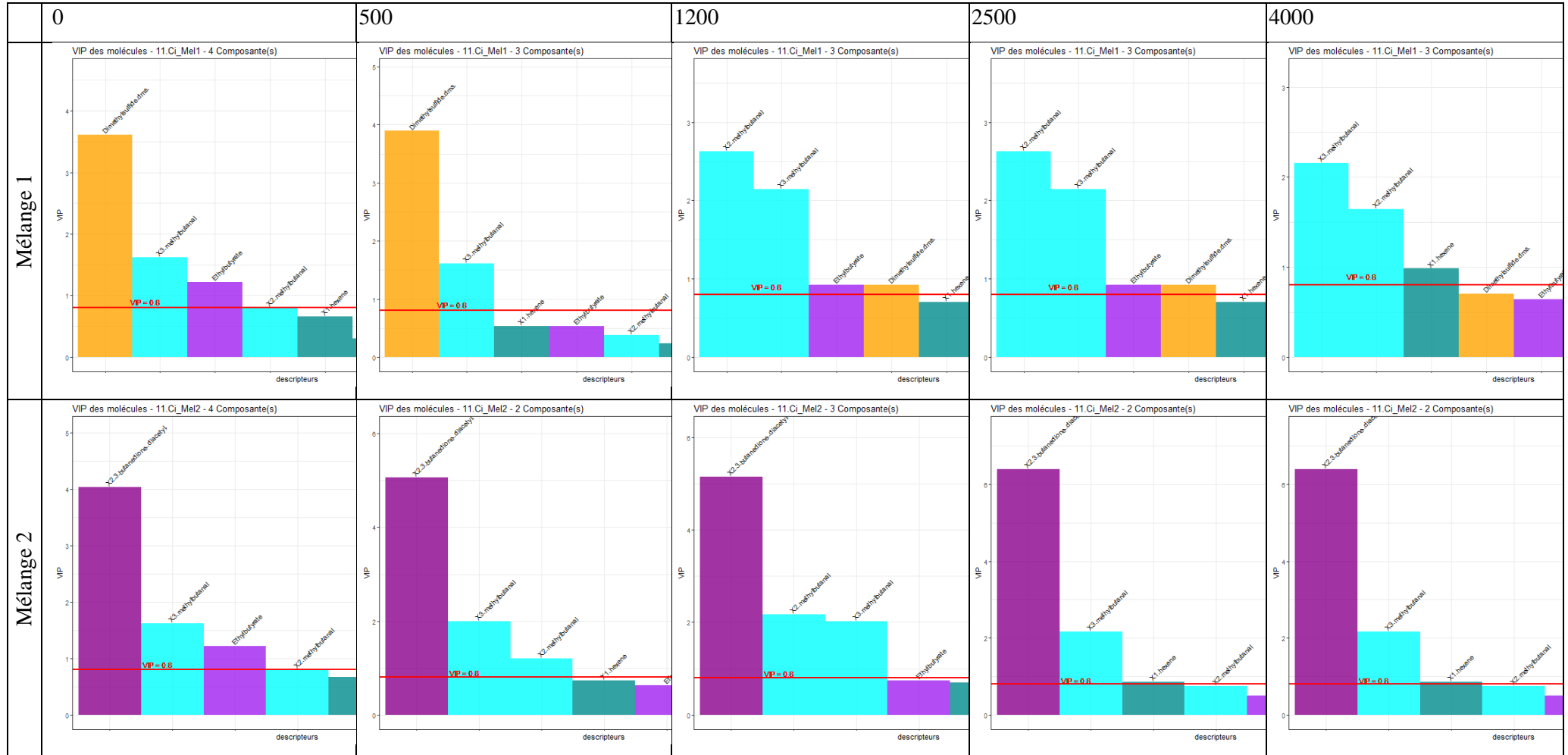
Annexe 23 COMPOSITION DE L'ESSAI LIQUIDE A

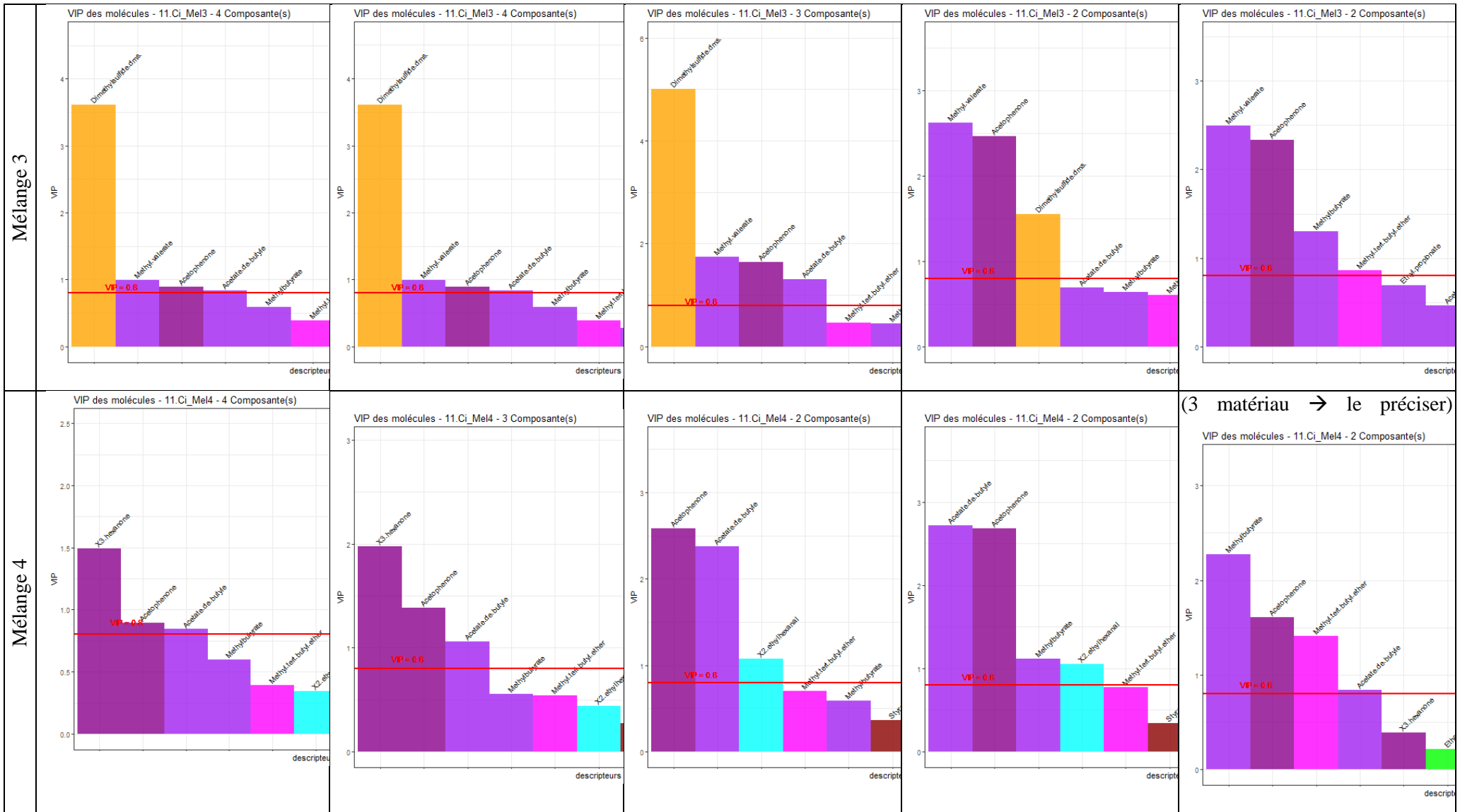
N° CAS	Molécule	Famille	Volume théorique (µL)	Densité	Masse théorique (mg)	Masse réelle (mg)	Volume réel (µL)
67-64-1	acetone	Cétones	100	0,79	79,0	69,4	87,8
109-66-0	pentane	Alcanes	100	0,649	64,9	45,5	70,1
60-29-7	diethylether	ethers	100	0,714	71,4	45,4	63,6
75-18-3	dimethylsulfide(DMS)	Soufrés	100	0,84	84,0	80,8	96,2
75-83-2	2,2-dimethylbutane	Alcanes	100	0,675	67,5	61,3	90,8
79-29-8	2, 3-dimethylbutane	Alcanes	100	0,662	66,2	57,2	86,4
1634-04-4	methyl tert-butyl ether	Ethers	100	0,75	75,0	61,1	81,5
96-14-0	3-methylpentane	Alcanes	100	0,673	67,3	61,4	91,2
78-93-3	2-butanone	11,10	100	0,786	78,6	76,2	96,9
592-41-6	1-hexene	Alcènes	100	0,678	67,8	62,5	92,2
110-54-3	hexane	Alcanes	100	0,659	65,9	60,4	91,7
141-78-6	acetate d'ethyle	Esters	100	0,898	89,8	86,4	96,2
109-99-9	tetrahydrofurane	Ethers	100	0,89	89,0	159,6	179,3
590-86-3	3-methylbutanal	Aldéhydes	100	0,791	79,1	77,3	97,7
71-36-3	1-butanol	Alcools	100	0,81	81,0	54,4	67,2
96-17-3	2-methylbutanal	Aldéhydes	100	0,791	79,1	82,0	103,7
110-83-8	cyclohexene	Alcènes	100	0,823	82,3	77,9	94,7
592-76-7	1-heptene	Alcènes	100	0,697	69,7	66,6	95,6
142-82-5	heptane	Alcanes	100	0,695	69,5	65,2	93,8
105-37-3	ethyl propionate	Esters	100	0,888	88,8	87,0	98,0
623-42-7	methylbutyrate	Esters	100	0,891	89,1	92,4	103,7
108-10-1	4-methyl-2-pentanone	Cétones	100	0,8	80,0	77,3	96,6
108-88-3	toluene	Aromatiques	100	0,871	87,1	85,6	98,3
589-38-8	3-hexanone	Cétones	100	0,803	80,3	78,2	97,4
591-78-6	2-hexanone	Cétones	100	0,81	81,0	77,6	95,8
111-66-0	1-octene	Alcènes	100	0,72	72,0	68,5	95,1
105-54-4	ethylbutyrate	Esters	100	0,886	88,6	86,1	97,2
111-65-9	octane	Alcanes	100	0,703	70,3	67,8	96,4
123-86-4	acetate de butyle	Esters	100	0,886	88,6	84,1	94,9
624-24-8	methyl valerate	Esters	100	0,886	88,6	87,8	99,1
100-41-4	ethylbenzene	Aromatiques	100	0,868	86,8	83,9	96,7
106-42-3	p-xylene	Aromatiques	100	0,86	86,0	81,6	94,9
108-38-3	m-xylene	Aromatiques	100	0,864	86,4	85,9	99,4
110-43-0	2-heptanone	Cétones	100	0,808	80,8	78,1	96,7
111-84-2	nonane	Alcanes	100	0,724	72,4	69,0	95,3
100-42-5	styrene	Aromatiques	100	0,902	90,2	86,7	96,1
95-47-6	o-xylene	Aromatiques	100	0,879	87,9	90,8	103,3
123-05-7	2-ethylhexanal	Aldéhydes	300	0,809	242,7	247,7	306,2
108-83-8	2,6-dimethyl-4-heptanone	Cétones	300	0,811	243,3	231,7	285,7
124-18-5	decane	Alcanes	300	0,735	220,5	218,0	296,6
95-63-6	1,2,4-trimethylbenzene	Aromatiques	300	0,88	264,0	254,7	289,4
104-76-7	2-ethylhexanol	Alcools	300	0,821	246,3	227,0	276,5
1120-21-4	undecane	Alcanes	300	0,743	222,9	214,3	288,4
98-86-2	acetophenone	Cétones	300	1,0266	308,0	311,6	303,5
112-31-2	decanal	Aldéhydes	300	0,83	249,0	243,1	292,9

Annexe 24 ANALYSE DE REPETABILITE DES ANALYSES PHYSICO-CHIMIQUES

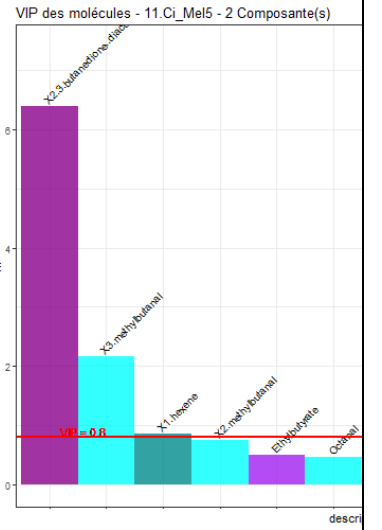
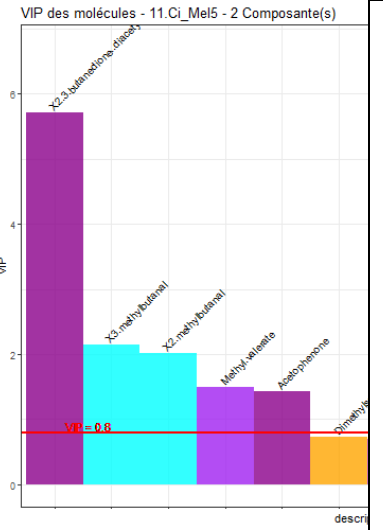
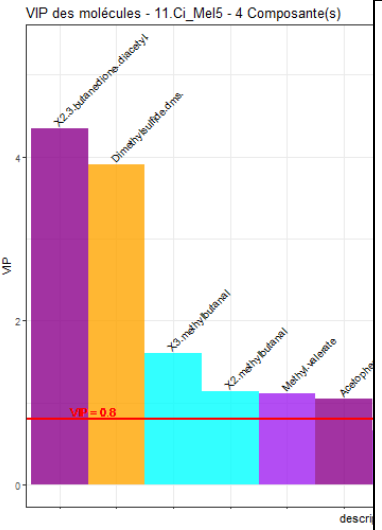
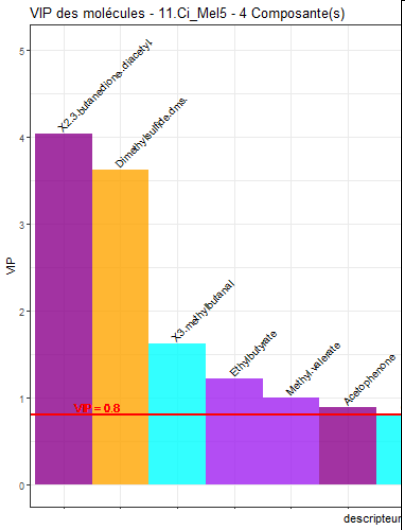
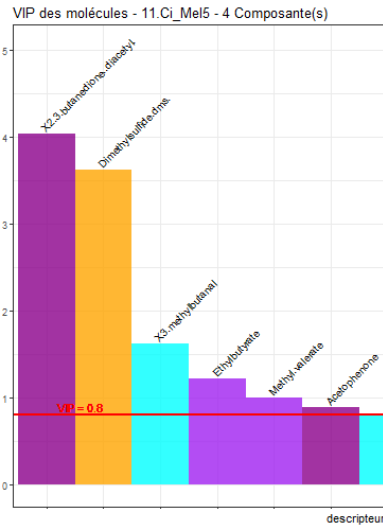
N° CAS	Molécule	Famille	Tube 2	Tube3	Tube 4	Tube 5	Moyenne	Ecart-type	RSD (%)
67-64-1	acetone	Cétones	609 512 000	596 053 043	628 085 259	598 472 994	608 030 824	12 641 139	2.08
109-66-0	pentane	Alcanes	787 529 708	827 001 809	842 454 333	849 970 080	826 738 983	24 104 421	2.92
60-29-7	diethylether	Ethers	580 539 736	638 370 439	624 477 233	650 483 733	623 467 785	26 437 601	4.24
75-18-3	DMS	Souffrés	929 975 756	965 398 713	951 138 410	945 582 445	948 023 831	12 681 488	1.34
75-83-2	2,2-dimethylbutane	Alcanes	1 324 128 351	1 297 020 039	1 319 395 182	1 302 295 014	1 310 709 647	11 332 603	0.86
79-29-8	2,3-dimethylbutane	Alcanes	1 133 240 299	1 164 113 486	1 184 658 143	1 213 618 481	1 173 907 602	29 335 483	2.50
1634-04-4	methyl tert-butyl ether	Ethers	1 035 582 457	1 113 456 396	1 082 149 462	1 070 881 560	1 075 517 469	27 837 216	2.59
96-14-0	3-methylpentane	Alcanes	1 333 880 412	1 357 265 786	1 376 450 168	1 325 718 952	1 348 328 830	19 941 428	1.48
78-93-3	2-butanone	Cétones	849 883 956	866 390 911	904 605 021	881 495 304	875 593 798	20 138 073	2.30
592-41-6	1-hexene	Alcènes	982 653 151	999 512 251	1 008 907 509	984 147 014	993 804 981	10 935 023	1.10
110-54-3	hexane	Alcanes	1 298 694 499	1 270 824 843	1 350 118 307	1 294 486 839	1 303 531 122	28 919 105	2.22
141-78-6	acetate d'ethyle	Esters	953 299 571	957 358 732	1 013 861 664	1 054 154 283	994 668 563	41 863 896	4.21
109-99-9	tetrahydrofurane	Ethers	1 082 625 352	1 125 288 154	1 169 455 649	1 141 043 245	1 129 603 100	31 403 145	2.78
590-86-3	3-methylbutanal	Aldéhydes	1 158 918 134	1 214 769 628	1 235 366 949	1 275 605 334	1 221 165 011	42 074 394	3.45
71-36-3	1-butanol	Alcools	644 944 261	698 085 783	707 539 156	713 811 973	691 095 293	27 226 976	3.94
96-17-3	2-methylbutanal	Aldéhydes	878 048 722	911 055 407	945 829 682	1 054 717 508	947 412 830	66 426 730	7.01
110-83-8	cyclohexene	Alcènes	1 498 756 324	1 586 823 113	1 637 161 817	1 603 600 826	1 581 585 520	51 140 753	3.23
592-76-7	1-heptene	Alcènes	1 302 713 344	1 383 999 260	1 445 284 949	1 399 662 146	1 382 914 925	51 487 762	3.72
142-82-5	heptane	Alcanes	1 431 023 602	1 536 157 889	1 576 401 959	1 578 385 646	1 530 492 274	59 848 956	3.91
105-37-3	ethyl propionate	Esters	1 114 974 428	1 188 838 066	1 237 689 518	1 232 097 098	1 193 399 778	49 067 322	4.11
623-42-7	methylbutyrate	Esters	1 182 475 117	1 282 495 684	1 332 081 845	1 307 758 579	1 276 202 806	56 883 024	4.46
108-10-1	4-methyl-2-pentanone	Cétones	1 208 293 185	1 316 982 579	1 363 361 405	1 334 775 271	1 305 853 110	58 705 833	4.50
108-88-3	toluene	Aromatiques	1 875 441 580	2 013 840 721	1 999 123 104	1 980 222 915	1 967 157 080	54 276 233	2.76
589-38-8	3-hexanone	Cétones	1 006 451 222	1 108 908 676	1 130 247 557	1 127 158 208	1 093 191 416	50 739 024	4.64
591-78-6	2-hexanone	Cétones	989 647 381	1 091 108 376	1 106 591 635	1 117 572 766	1 076 230 040	50 864 942	4.73
111-66-0	1-octene	Alcènes	1 345 696 702	1 475 304 928	1 464 418 556	1 454 539 691	1 434 989 969	52 074 026	3.63
105-54-4	ethylbutyrate	Esters	1 187 161 668	1 298 936 950	1 311 843 106	1 291 247 833	1 272 297 389	49 700 956	3.91
111-65-9	octane	Alcanes	1 436 693 257	1 609 809 177	1 611 589 943	1 578 853 920	1 559 236 574	71 937 765	4.61
123-86-4	acetate de butyle	Esters	1 109 695 909	1 220 549 736	1 240 626 559	1 231 105 717	1 200 494 480	52 901 383	4.41
624-24-8	methyl valerate	Esters	1 195 430 288	1 312 058 513	1 331 037 374	1 317 541 213	1 289 016 847	54 471 843	4.23
100-41-4	ethylbenzene	Aromatiques	1 775 429 805	1 894 067 665	1 919 197 211	1 893 813 811	1 870 627 123	55 921 070	2.99
106-42-3	p-xylene	Aromatiques	1 816 496 625	1 842 041 730	1 855 649 372	1 953 341 133	1 866 882 215	51 858 169	2.78
108-38-3	m-xylene	Aromatiques	1 635 600 997	1 815 501 646	1 891 925 365	1 752 694 931	1 773 930 735	93 857 537	5.29
110-43-0	2-heptanone	Cétones	981 503 953	1 102 750 205	1 144 540 206	1 143 996 120	1 093 197 621	66 676 975	6.10
111-84-2	nonane	Alcanes	1 304 030 614	1 409 386 929	1 450 619 957	1 438 435 360	1 400 618 215	57 741 592	4.12
100-42-5	styrene	Aromatiques	1 302 381 869	1 368 479 502	1 407 397 418	1 451 268 180	1 382 381 742	54 690 898	3.96
95-47-6	o-xylene	Aromatiques	1 950 754 323	2 020 388 778	2 093 389 214	2 127 314 658	2 047 961 743	68 136 061	3.33
123-05-7	2-ethylhexanal	Aldéhydes	3 831 474 612	4 004 125 749	4 211 128 998	4 112 702 560	4 039 857 980	140 837 242	3.49
108-83-8	2,6-dimethyl-4-heptanone	Cétones	2 875 798 270	3 061 558 774	3 248 893 717	3 185 005 740	3 092 814 125	142 243 551	4.60
124-18-5	decane	Alcanes	3 621 116 880	3 787 369 717	4 071 271 240	3 889 860 960	3 842 404 699	163 270 587	4.25
95-63-6	1,2,4-trimethylbenzene	Aromatiques	4 590 442 187	4 944 491 367	5 261 902 895	5 108 060 883	4 976 224 333	249 413 293	5.01
104-76-7	2-ethylhexanol	Alcools	2 286 928 439	2 447 785 017	2 647 542 852	2 546 144 378	2 482 100 172	132 987 289	5.36
1120-21-4	undecane	Alcanes	3 995 965 935	4 160 089 906	4 379 183 204	4 150 964 796	4 171 550 960	136 470 139	3.27
98-86-2	acetophenone	Cétones	788 815 762	948 987 944	1 128 740 916	967 744 032	958 572 164	120 364 676	13
112-31-2	decanal	Aldéhydes	1 632 351 644	1 782 541 578	1 907 187 295	1 735 038 645	1 764 279 791	98 762 503	6

Annexe 25 RESULTATS PLS PAR MELANGE SUIVANT LE SEUIL DE SIMILITUDE UTILISE





Mélange 5



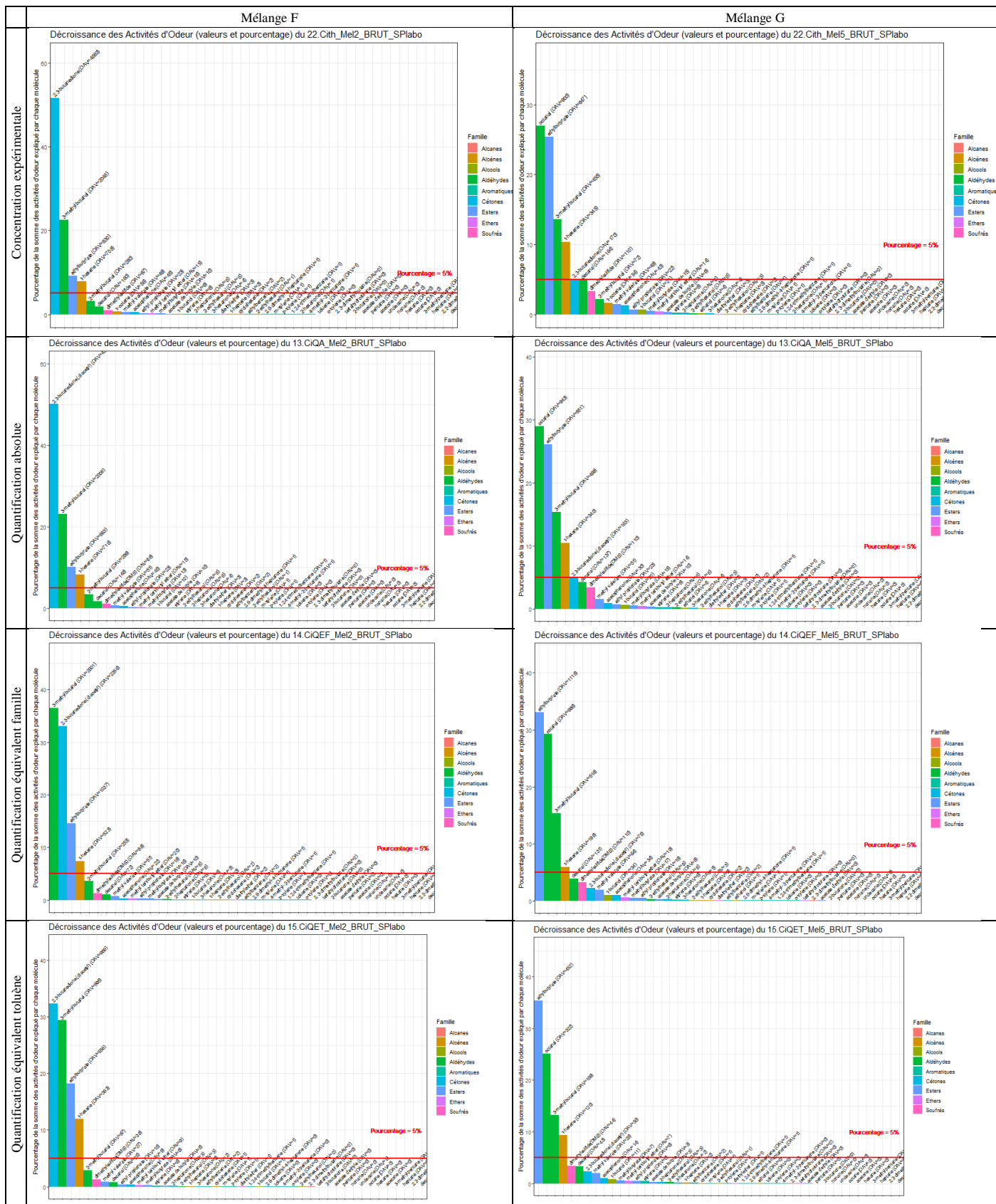
**Annexe 26 QUANTIFICATION DES MOLECULES DU MELANGE F A PARTIR DES
COURANTS IONIQUES TOTAUX PAR LES DIFFERENTES METHODES**

N° CAS	Molécule	Famille	$C_{i_{th}}$ (mg/m ³)	$C_{i_{abs}}$ (mg/m ³)	$C_{i_{eq,fam}}$ (mg éq.fam/m ³)	$C_{i_{eq,tol}}$ (mg éq.tol/m ³)
109-66-0	pentane	Alcanes	0.24	0.25	0.24	0.19
79-29-8	2, 3-dimethylbutane	Alcanes	0.41	0.39	0.39	0.32
75-83-2	2,2-dimethylbutane	Alcanes	0.31	0.31	0.30	0.25
96-14-0	3-methylpentane	Alcanes	0.36	0.36	0.34	0.28
110-54-3	hexane	Alcanes	0.42	0.42	0.41	0.33
142-82-5	heptane	Alcanes	0.38	0.38	0.42	0.35
111-65-9	octane	Alcanes	0.39	0.38	0.42	0.35
111-84-2	nonane	Alcanes	0.42	0.44	0.50	0.40
124-18-5	decane	Alcanes	0.40	0.42	0.40	0.32
1120-21-4	undecane	Alcanes	0.41	0.42	0.36	0.29
592-41-6	1-hexene	Alcènes	0.37	0.37	0.27	0.19
110-83-8	cyclohexene	Alcènes	0.43	0.43	0.50	0.34
592-76-7	1-heptene	Alcènes	0.39	0.39	0.43	0.30
71-36-3	1-butanol	Alcools	0.46	0.52	0.30	0.03
104-76-7	2-ethylhexanol	Alcools	0.43	0.47	0.88	0.10
590-86-3	3-methylbutanal	Aldéhydes	0.45	0.44	0.57	0.20
96-17-3	2-methylbutanal	Aldéhydes	0.47	0.50	0.42	0.15
123-05-7	2-ethylhexanal	Aldéhydes	0.45	0.47	0.65	0.22
112-31-2	decanal	Aldéhydes	0.45	0.41	0.20	0.07
108-88-3	toluene	Aromatiques	0.49	0.50	0.53	0.50
100-41-4	ethylbenzene	Aromatiques	0.48	0.49	0.53	0.50
106-42-3	p-xylene	Aromatiques	0.48	0.49	0.55	0.52
108-38-3	m-xylene	Aromatiques	0.48	0.50	0.49	0.47
100-42-5	styrene	Aromatiques	0.50	0.50	0.38	0.35
95-47-6	o-xylene	Aromatiques	0.49	0.49	0.54	0.51
95-63-6	1,2,4-trimethylbenzene	Aromatiques	0.49	0.49	0.44	0.42
67-64-1	acetone	Cétones	0.40	0.42	0.41	0.17
431-03-8	2,3-butanedione	Cétones	0.55	0.51	0.28	0.12
78-93-3	2-butanone	Cétones	0.44	0.47	0.51	0.21
108-10-1	4-methyl-2-pentanone	Cétones	0.44	0.45	0.69	0.29
589-38-8	3-hexanone	Cétones	0.45	0.45	0.52	0.22
110-43-0	2-heptanone	Cétones	0.46	0.47	0.61	0.26
108-83-8	2,6-dimethyl-4-heptanone	Cétones	0.45	0.50	0.47	0.20
98-86-2	acetophenone	Cétones	0.58	0.58	0.29	0.12
141-78-6	acetate d'ethyle	Esters	0.50	0.51	0.36	0.19
105-37-3	ethyl propionate	Esters	0.50	0.50	0.40	0.21
623-42-7	methylbutyrate	Esters	0.50	0.43	0.54	0.29
105-54-4	ethylbutyrate	Esters	0.48	0.51	0.60	0.32
123-86-4	acetate de butyle	Esters	0.49	0.50	0.52	0.28
624-24-8	methyl valerate	Esters	0.49	0.51	0.51	0.27
60-29-7	diethylether	Ethers	0.28	0.28	0.17	0.08
1634-04-4	methyl tert-butyl ether	Ethers	0.38	0.38	0.50	0.23
109-99-9	tetrahydrofurane	Ethers	0.47	0.48	0.51	0.23
75-18-3	dimethylsulfide	Soufrés	0.35	0.35	0.35	0.15
111-66-0	1-octene	Alcènes	0.40	co-élution avec 2-hexanone		
591-78-6	2-hexanone	Cétones	0.45	co-élution avec 1-octene		

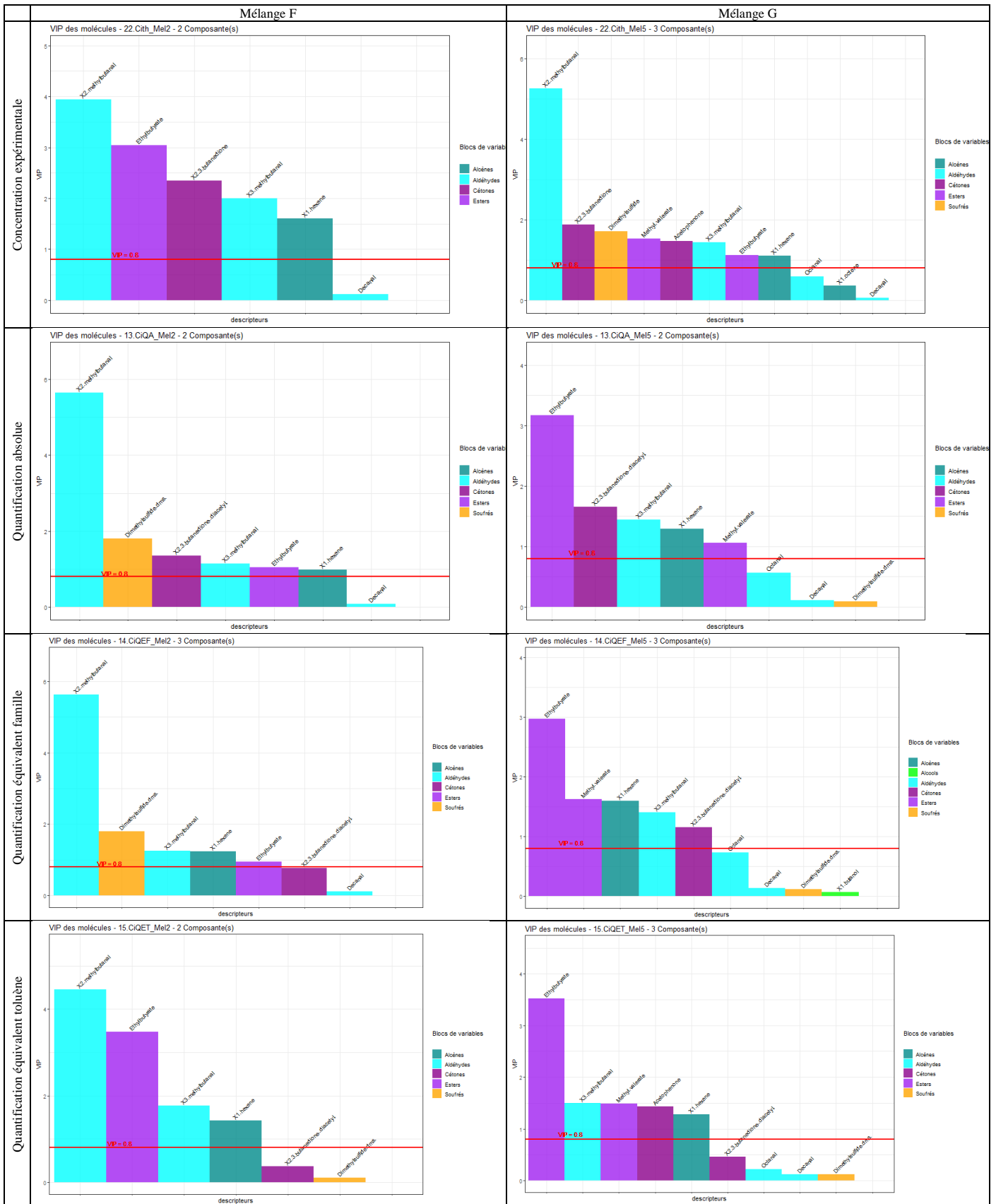
Annexe 27 QUANTIFICATION DES MOLECULES DU MELANGE G A PARTIR DES COURANTS IONIQUES TOTAUX PAR LES DIFFERENTES METHODES

N° CAS	Molécules	Famille	$C_{i_{th}}$ (mg/m ³)	$C_{i_{abs}}$ (mg/m ³)	$C_{i_{eq,fam}}$ (mg éq. fam/m ³)	$C_{i_{eq,tol}}$ (mg éq. tol/m ³) 3)
109-66-0	pentane	Alcanes	0.30	0.29	0.30	0.21
75-83-2	2,2-dimethylbutane	Alcanes	0.29	0.28	0.26	0.18
79-29-8	2, 3-dimethylbutane	Alcanes	0.31	0.30	0.27	0.19
96-14-0	3-methylpentane	Alcanes	0.33	0.33	0.29	0.21
110-54-3	hexane	Alcanes	0.34	0.34	0.28	0.20
142-82-5	heptane	Alcanes	0.37	0.38	0.43	0.30
111-65-9	octane	Alcanes	0.29	0.29	0.35	0.25
111-84-2	nonane	Alcanes	0.39	0.41	0.48	0.34
124-18-5	decane	Alcanes	0.42	0.43	0.41	0.29
1120-21-4	undecane	Alcanes	0.42	0.44	0.38	0.27
592-41-6	1-hexene	Alcènes	0.18	0.18	0.10	0.06
110-83-8	cyclohexene	Alcènes	0.43	0.43	0.51	0.31
592-76-7	1-heptene	Alcènes	0.37	0.38	0.42	0.25
64-17-5	ethanol	Alcools	0.42	0.41	0.13	0.04
71-36-3	1-butanol	Alcools	1.10	1.06	1.79	0.56
104-76-7	2-ethylhexanol	Alcools	1.09	1.18	1.16	0.36
590-86-3	3-methylbutanal	Aldéhydes	0.10	0.11	0.11	0.04
123-05-7	2-ethylhexanal	Aldéhydes	0.45	0.48	0.72	0.24
124-13-0	octanal	Aldéhydes	0.45	0.47	0.49	0.16
112-31-2	decanal	Aldéhydes	0.46	0.38	0.37	0.12
108-88-3	toluene	Aromatiques	0.48	0.50	0.54	0.50
100-41-4	ethylbenzene	Aromatiques	0.47	0.50	0.52	0.48
106-42-3	p-xylene	Aromatiques	0.48	0.50	0.49	0.45
108-38-3	m-xylene	Aromatiques	0.47	0.48	0.55	0.51
100-42-5	styrene	Aromatiques	0.49	0.49	0.37	0.35
95-47-6	o-xylene	Aromatiques	0.49	0.51	0.57	0.53
95-63-6	1,2,4-trimethylbenzene	Aromatiques	0.47	0.48	0.42	0.39
67-64-1	acetone	Cétones	0.40	0.40	0.32	0.13
431-03-8	2,3-butanedione	Cétones	0.02	0.02	0.01	0.00
78-93-3	2-butanone	Cétones	0.43	0.43	0.30	0.12
108-10-1	4-methyl-2-pentanone	Cétones	0.43	0.49	0.56	0.23
589-38-8	3-hexanone	Cétones	0.48	0.43	0.53	0.22
110-43-0	2-heptanone	Cétones	0.47	0.50	0.60	0.24
108-83-8	2,6-dimethyl-4-heptanone	Cétones	0.43	0.46	0.51	0.21
98-86-2	acetophenone	Cétones	0.56	0.39	0.43	0.18
141-78-6	acetate d'ethyle	Esters	0.49	0.50	0.41	0.16
105-37-3	ethyl propionate	Esters	0.49	0.50	0.34	0.14
623-42-7	methylbutyrate	Esters	0.51	0.51	0.57	0.23
105-54-4	ethylbutyrate	Esters	0.49	0.49	0.65	0.26
123-86-4	acetate de butyle	Esters	0.48	0.51	0.42	0.17
624-24-8	methyl valerate	Esters	0.49	0.51	0.64	0.26
60-29-7	diethylether	Ethers	0.30	0.30	0.18	0.07
1634-04-4	methyl tert-butyl ether	Ethers	0.35	0.35	0.46	0.17
109-99-9	tetrahydrofurane	Ethers	0.47	0.47	0.51	0.19
75-18-3	dimethylsulfide	Soufrés	0.44	0.44	0.44	0.18
111-66-0	1-octene	Alcènes	0.39	co-élué avec 2-hexanone		
591-78-6	2-hexanone	Cétones	0.43	co-élué avec 1-octène		
96-17-3	2-methylbutanal	Aldéhydes	0.12	co-élué avec butanol		

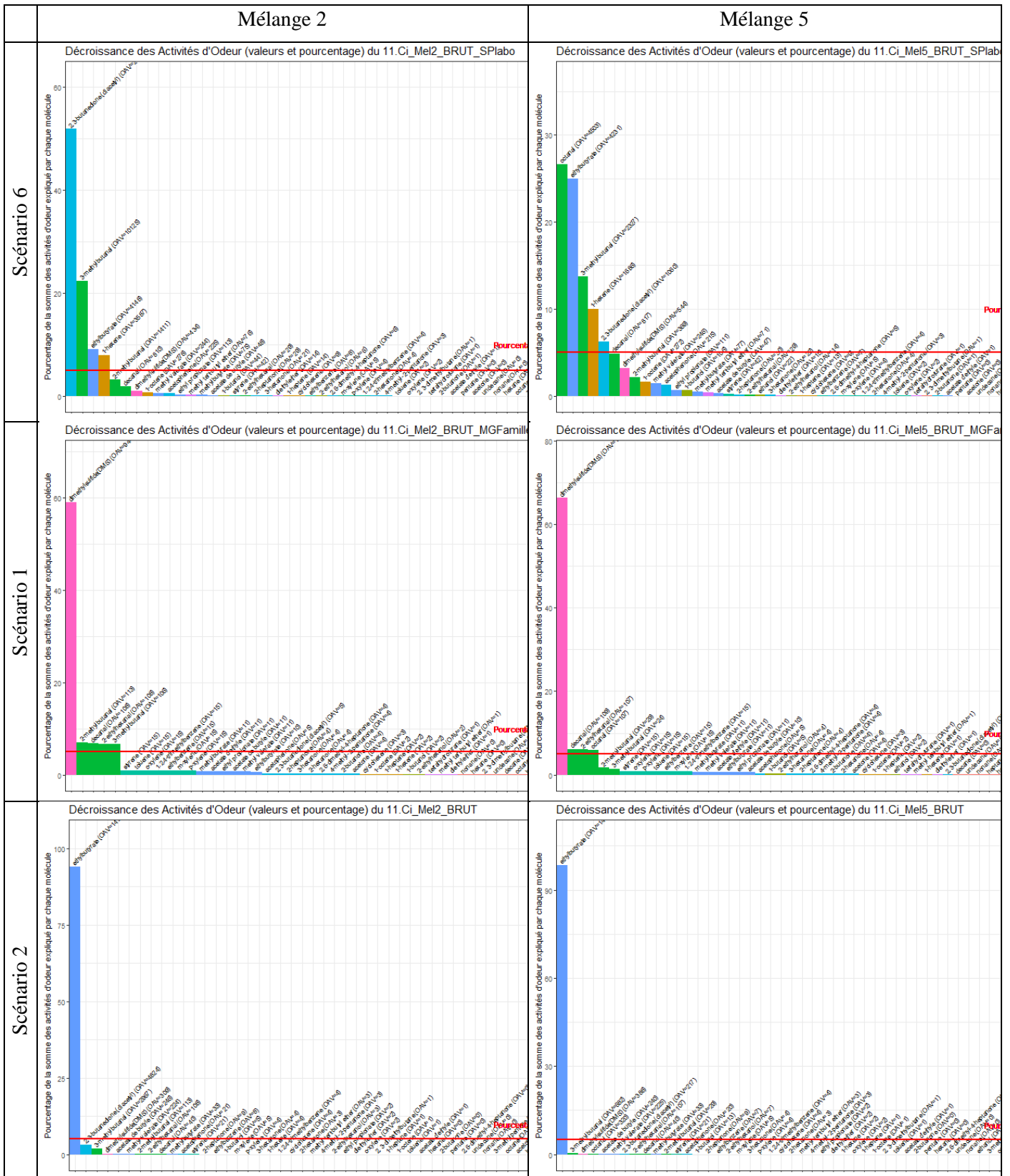
Annexe 28 GRAPHIQUE DES OAV DES MELANGES F ET G EN FONCTION DE LA QUANTIFICATION APPLIQUEE

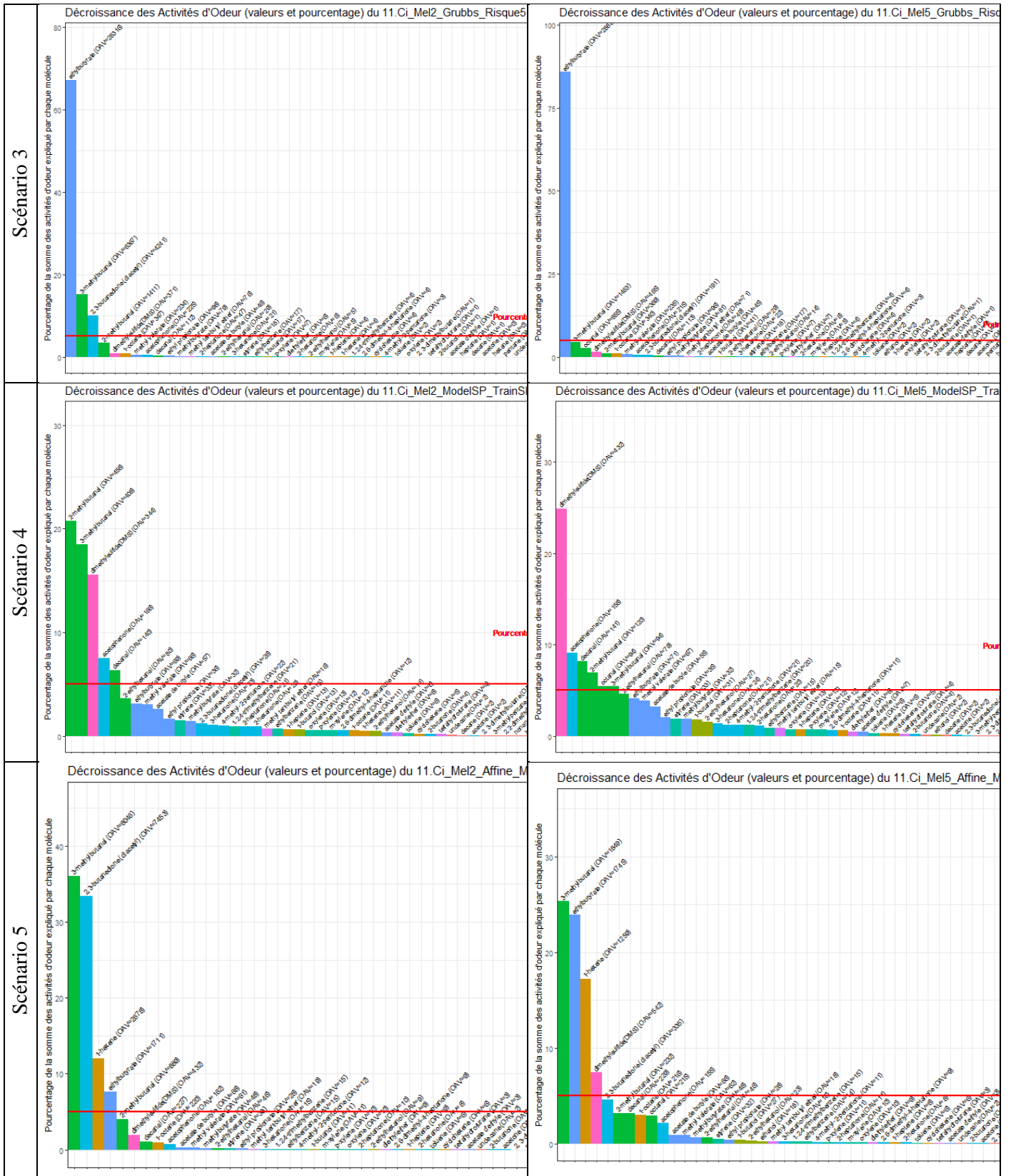


Annexe 29 GRAPHES DES VIP RESULTANT DES PLS APPLIQUEES AUX ETUDES DES MELANGES F ET G EN FONCTION DE LA QUANTIFICATION APPLIQUEE

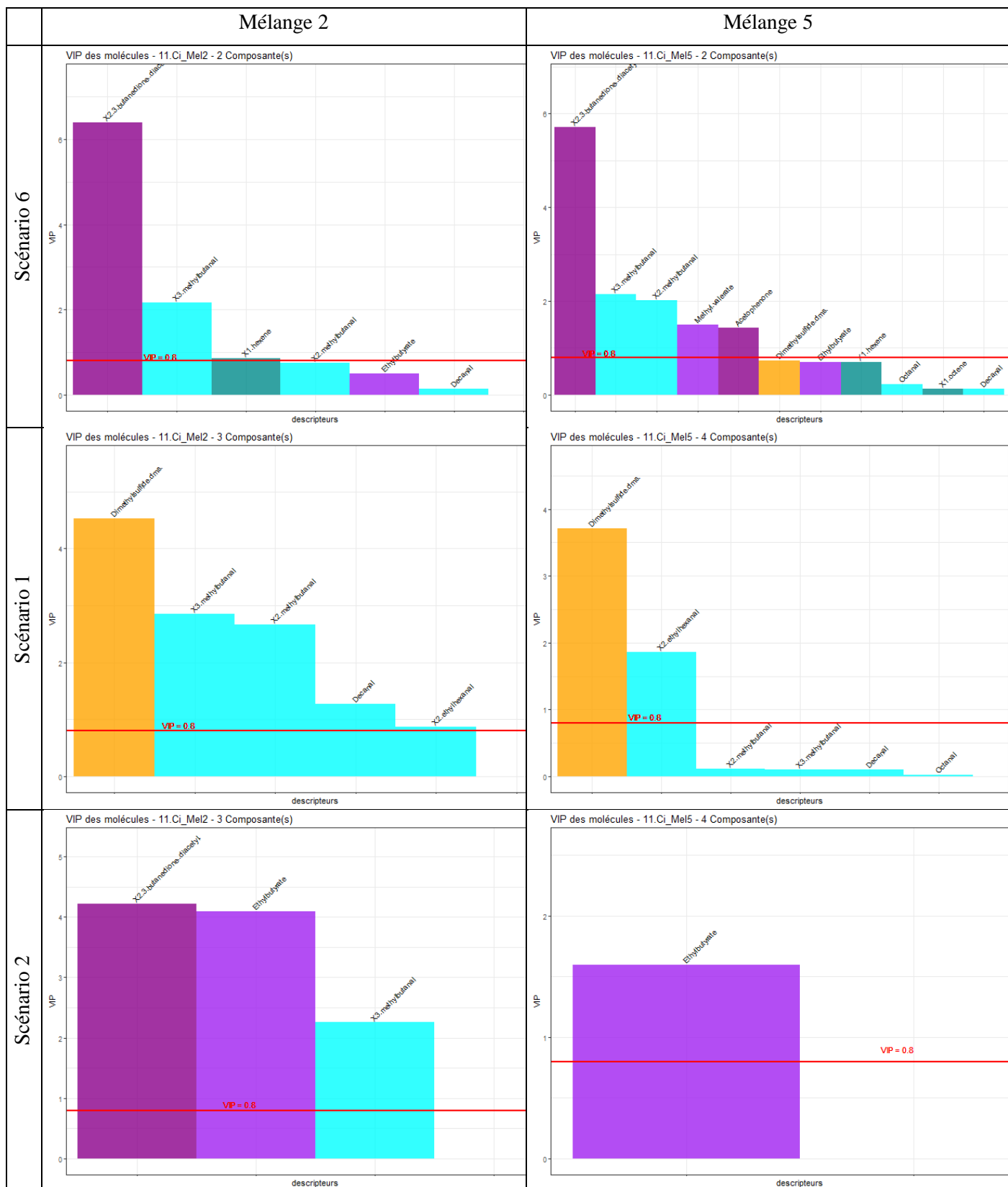


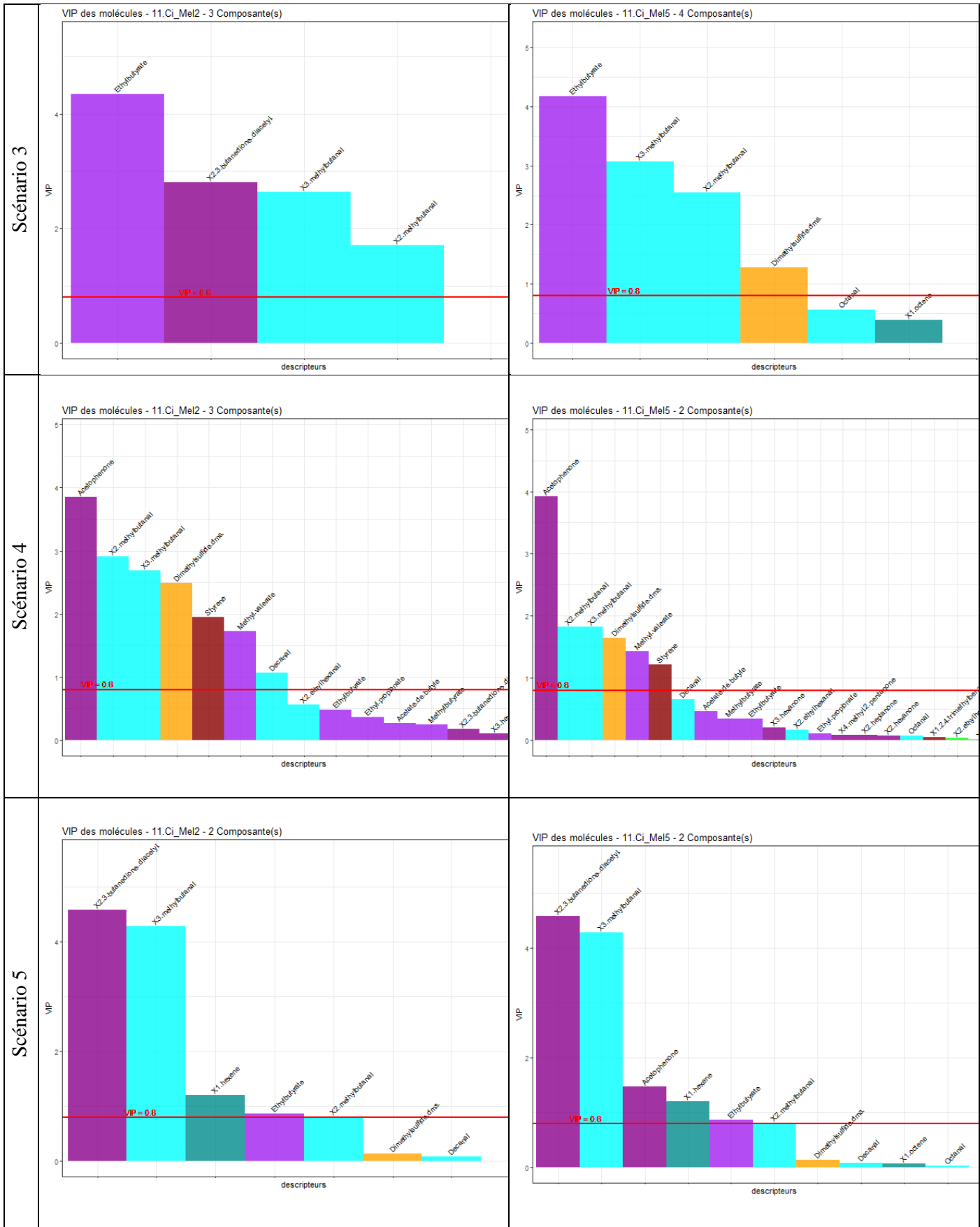
Annexe 30 GRAPHS DES OAV DES MELANGES 2 ET 5 EN FONCTION DES SCENARIOS DE SEUILS DE PERCEPTION APPLIQUES



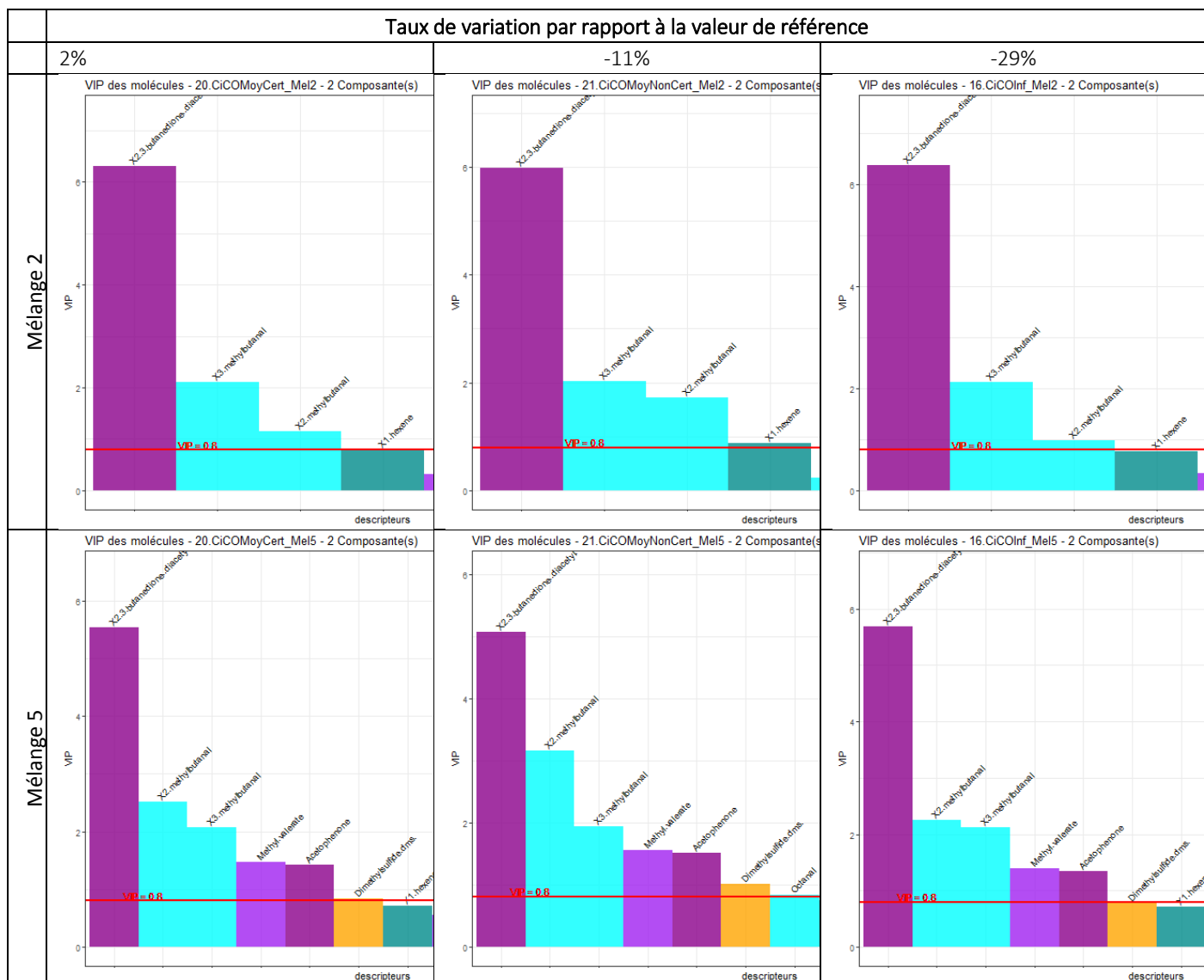


Annexe 31 GRAPHS DES VIP RESULTANT DES PLS APPLIQUEES AUX ETUDES DES MELANGES 2 ET 5 EN FONCTION DES SCENARIOS DE SEUILS DE PERCEPTION APPLIQUES





Annexe 32 GRAPHS DES VIP RESULTANT DES PLS APPLIQUEES AUX ETUDES DES MELANGES 2 ET 5 EN FONCTION DU TAUX DE VARIATION DE LA CONCENTRATION D'ODEUR EXPERIMENTALE PAR RAPPORT A LA REFERENCE DU JURY 1



Annexe 33 ORGANIGRAMME RESUME DE LA DEMARCHE

