

THESE

Pour obtenir le grade de
Docteur de l'Université de Bourgogne

Discipline :
Sciences de l'Alimentation

par

Caroline Peltier

L'analyse statistique du profil sensoriel revisitée par une approche base de données

Directeur de thèse

Pascal Schlich

Co-encadrant

Michel Visalli

Jury

M. CARDOT Hervé

M. MAC FIE Hal

Mme VIGNEAU Evelyne

M. HUSSON François

Mme SAINT-EVE Anne

M. SCHLICH Pascal

M. VISALLI Michel

Professeur, Université de Bourgogne, France

Professeur associé, Reading University, UK

Professeur, Oniris, Nantes, France

Professeur, Agrocampus Ouest, Rennes, France

Docteur, AgroParisTech, France

Docteur, INRA, CSGA, Dijon, France

Ingénieur, INRA, CSGA, Dijon, France

Président

Rapporteur

Rapporteur

Examinateur

Examinateur

Directeur de thèse

Co-encadrant de thèse

AVANT-PROPOS

Cette thèse a été réalisée dans le cadre d'une bourse Jeune Chercheur Entrepreneur accordée par la Région Bourgogne avec le soutien de 6 parrains industriels : Biofortis, Kraft Food (Mondelez), Nestlé, OPP, Pernod Ricard, RMT Sensorialis.

REMERCIEMENTS

Je remercie la région Bourgogne pour avoir financé cette thèse par le dispositif Jeune Chercheur Entrepreneur.

Je tiens ensuite à remercier mon directeur de thèse, Pascal Schlich, qui m'a permis de me lancer dans ce travail de thèse et dont les remarques et les idées m'ont permis de progresser durant ces trois ans.

Je remercie évidemment mon co-encadrant, Michel Visalli, pour son efficacité, sa disponibilité et son soutien.

Merci aux membres de mon comité de thèse d'avoir accepté de me suivre pendant ces trois ans : Hervé Cardot et El Mostapha Qannari.

Merci aux représentants des différents parrains industriels de cette thèse dont les remarques ont apporté un éclairage précieux sur les besoins des industriels : Kevin Guillamet, Geneviève Guillot, Virginie Herbreteau, Sébastien Leuillet, Marie-Christine Marcuz, Nicolas Pineau, Pieter Punter, David Viala.

Merci au jury de ma thèse. Un merci tout particulier à Hervé Cardot pour avoir accepté de présider ce jury.

Merci aux rapporteurs d'avoir accepté de rapporter sur ce manuscrit pendant leurs vacances d'été : Evelyne Vigneau et Hal Mac Fie. Merci pour vos remarques.

Merci aux examinateurs de mon jury de thèse d'avoir accepté ce rôle: Anne Saint-Eve et François Husson.

Merci à Per Brockhoff et son équipe danoise qui m'ont permis de passer quelque temps dans leur laboratoire à l'université de Lyngby.

Merci à tous les membres de la plateforme ChemoSens pour leur bonne humeur.

Merci aux relecteurs (trices) de cette thèse.

Et de manière plus personnelle, un immense merci à tous ceux et celles qui ont compté et m'ont accompagnée pendant ces trois ans.

Résumé

Lors d'une épreuve d'analyse sensorielle, des sujets préalablement entraînés dégustent des produits et les notent à l'aide de différents descripteurs afin d'établir des profils descriptifs. De nombreuses méthodes statistiques (ACP, ANOVA...) permettent d'analyser les données obtenues, afin notamment d'obtenir des résultats sur les performances des sujets, et de représenter les produits sur des cartes.

Durant ce travail de thèse, ces méthodes statistiques ont été revisitées par l'exploitation d'un grand nombre de jeux de données issus de la SensoBase (base de données contenant plus de 1000 jeux de données de profil sensoriel). Différentes méthodes de cartographie et d'analyse des performances des sujets ont ainsi été comparées, puis des améliorations ont été proposées. La prise en compte du scaling (tendance d'un sujet à disperser ses notes plus –ou moins- que le panel) et l'adaptation de cartes à des modèles statistiques spécifiques ont été tout particulièrement étudiées.

Ce travail a abouti à la documentation de certaines pratiques de l'analyse sensorielle et à l'établissement de recommandations à destination des animateurs de panel. Les différentes méthodes proposées dans cette thèse ont été implémentées sous R (bibliothèques CVAPack, MAMCAP, MultiMAM), et sont donc directement utilisables par tout animateur de panel. Tous ces outils sont transposables dans d'autres domaines scientifiques utilisant l'être humain comme instrument de mesure (psychologie, économie..).

Abstract

During a sensory evaluation, trained subjects taste and score products on different descriptors in order to obtain a descriptive profile of these products. The data are then analyzed by several statistical methods (PCA, ANOVA...) in order to monitor the subject performances and map the product space.

This work aims to revisit these statistical methods thanks to the analysis of a large number of datasets of sensory profiling data from the SensoBase (a database containing more than a thousand of datasets). Different statistical methods of mapping and analysis of performances were compared, then improved. We mainly focused on the so-called scaling effect (tendency of the subject to spread his scores more or less than the panel) and on maps related to specific statistical models.

This works led to some findings about sensory analysis and recommendations for panel leaders. Furthermore, different statistical tools (R packages such as CVAPack, MAMCAP and MultiMAM) were offered. These tools could also be useful in other scientific fields using human beings as a measurement instrument (psychology, economics, marketing...).

VALORISATION DES TRAVAUX

Publications

Peltier, C., Mammasse, N., Visalli, M., Cordelle, S., Schlich, P. (2015). Do we need to replicate in profile sensory studies? *Submitted to Food Quality and Preference*.

Peltier, C., Visalli, M., Schlich, P. (2015). Taking scaling effect into account in multidimensional analysis with the MultiMAM package. *Submitted to Journal of Statistical Software*.

Peltier, C., Visalli, M., Schlich, P. (2015). Multiplicative decomposition of the scaling effect in the Mixed Assessor Model into a descriptor-specific and an overall coefficients. *In review in Food Quality and Preference*.

Peltier, C., Visalli, M., Schlich, P. (2015). Canonical Variate Analysis of sensory profiling data. *In review in Journal Of Sensory Studies*.

Peltier, C., Visalli, M., Schlich, P. (2014). Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies. *Food Quality And Preference, 40B, 326-333*.

Peltier, C., Brockhoff, P.B., Visalli, M., Schlich, P. (2014). The MAM-CAP table: a new tool to monitor panel performances. *Food Quality and Preference, 32A 24-27*.

Communications orales

Peltier, C., Visalli, M., Schlich, P. (2014). Multidimensional extension of the Mixed Assessor Model. *In Sensometrics, Chicago July 29th to August 1st*.

Peltier, C., Visalli, M., Schlich, P. (2014). Multidimensional extension of scaling. Meta-analysis of 187 datasets. *In Agrostats, Rabat, Maroc, March 24-28*.

Peltier, C., Visalli, M., Schlich, P., (2014). A propos de la dispersion des notes des panélistes: Concepts et méta-analyses. *A la journée des Doctorants, Centre des Sciences du Goût et de l'Alimentation, Dijon, le 19 Juin*.

Peltier, C., Visalli, M., Schlich, P. (2013). Analyse en composantes principales vs. analyse en variables canoniques (CVA) : comparaison sur 422 jeux de données. *A la journée des Doctorants, Centre des Sciences du Goût et de l'Alimentation, Dijon, le 25 Juin*.

Peltier, C., Visalli, M., Schlich, P. (2013). Analyse en composantes principales vs. analyse en variables canoniques (CVA) : comparaison sur 422 jeux de données. *Au Forum des Jeunes Chercheurs, Dijon, le 13 Juin.*

Peltier, C., Visalli, M., Schlich, P. (2013). PCA vs. CVA: Meta-analysis of 422 datasets. In Pangborn, Rio, Brazil, August, 11-15.

Peltier, C., Visalli, M., Schlich, P. (2012). Monitoring panel performances with the Mixed Assessor Model . Meta-analysis of the SensoBase. In Sensometrics, Rennes, France, July 10-13.

Posters

Peltier, C., Visalli, M., Schlich, P. (2015). Should I replace replicates by additional panelists in my trained descriptive sensory panel ? In Pangborn, Gothenburg, Sweden, August 23-27.

Peltier, C., Pineau, N., Visalli, M., Schlich, P. (2015). The effect of gender, age and smoking status on use of scale, repeatability, product discrimination and agreement with the panel investigated in 177 descriptive sensory studies. In Pangborn, Gothenburg, Sweden, August 23-27.

Peltier, C., Visalli, M., Schlich, P. (2015). The subject effect in descriptive analysis : fixed or random? An old question revisited thanks to 697 datasets from the Sensobase. In Pangborn, Gothenburg, Sweden, August 23-27.

Librairies R mises à disposition:

CVAPack : disponible sur www.timesens.com le 03/07/2015

MAMCAP : disponible sur www.timesens.com le 01/07/2014

MultiMAM : disponible sur www.timesens.com, le 03/07/2015

TABLE DES MATIERES

Préambule	1
Chapitre 1: Introduction	3
A. Analyse sensorielle	5
I. Présentation de l'analyse sensorielle	5
II. L'être humain comme outil de mesure	5
1. L'être humain, un instrument de mesure des cinq sens.....	5
2. L'être humain : un instrument de mesure soumis à différents biais.....	7
III. Mise en place d'une étude sensorielle	10
IV. TimeSens, un logiciel d'acquisition et d'analyse des données d'analyse sensorielle ..	10
B. Analyse descriptive quantitative (QDA).....	12
I. Définition et historique des analyses descriptives	12
1. Définition.....	12
2. Avant le protocole QDA : historique de l'analyse descriptive	12
3. La méthode QDA	12
4. Autres protocoles issus de la méthode QDA.....	13
II. Questions classiques sur le protocole de profil sensoriel	14
1. Entraînement des sujets.....	14
2. Nombre de sujets, produits, descripteurs, répétitions.....	14
III. Analyses classiques des données de profil sensoriel	15
1. Performances	15
2. Cartes sensorielles.....	16
3. Différences entre produits	16
IV. Revue bibliographique des études QDA réalisées en 2014	17
C. Eléments théoriques sur les analyses statistiques classiques	20
I. Notations	20
II. Analyse en Composantes Principales (ACP)	21
1. L'ACP du tableau des moyennes des produits	21
2. Biplot	23
III. ANOVA et détection de différences unidimensionnelles entre les produits	25
1. Principe.....	25

2.	Théorie	25
3.	Exemple	28
4.	Interprétation	28
IV.	Problèmes théoriques de l'ACP et de l'ANOVA.....	28
D.	Objectifs et plan de cette thèse	29
Chapitre 2 :	Matériel et méthodes.....	33
A.	SensoBase	35
I.	Présentation	35
1.	Définition et historique	35
2.	Structure de 2006 à 2015	35
3.	Evolution de la SensoBase.....	35
4.	Sous ensemble de données utilisé	36
II.	Description de la SensoBase	36
1.	Caractéristiques des jeux de données.....	36
2.	Caractéristiques des sujets.....	38
3.	Caractéristiques des produits de la SensoBase.....	41
4.	Caractéristiques des descripteurs de la SensoBase	42
5.	Caractéristiques des fournisseurs	43
6.	Jeux de données non équilibrés.....	43
7.	Synthèse	44
B.	Méta-analyses.....	45
I.	Définition et introduction	45
II.	Historique des méta-analyses	45
III.	Méthodologie pour les méta-analyses sur essais cliniques	46
IV.	Méta-analyses déjà effectuées sur les bases de données sensorielles PrefBase et SensoBase	49
1.	Méta-analyse de Pineau (2006)	49
2.	Méta-analyse de Mammasse (2012).....	50
3.	Méta-analyse de Brockhoff (2015).....	50
C.	Méthodologie pour les méta-analyses en profil sensoriel	51
I.	Présentation de la méthodologie établie	51
II.	Outils disponibles pour la méta-analyse	52
D.	Exemple : utilisation de la méthodologie des méta-analyses sur SensoBase	54
I.	Etude des zéros dans la SensoBase	54

1.	Problématique.....	54
2.	Protocole	54
3.	Caractérisation de la base de données obtenue.....	54
4.	Résultats	54
5.	Commentaires	55
II.	Etude de l'impact du type d'effet sujet (fixe ou aléatoire) dans l'ANOVA.....	56
1.	Problématique.....	56
2.	Protocole	56
3.	Caractérisation de la base de données	56
4.	Résultats et interprétation	57
E.	Conclusion	58
Chapitre 3 : Nouvelles méthodes d'analyse statistique des données de profil sensoriel ...		59
A.	Contrôle des performances	61
I.	Performances unidimensionnelles	61
1.	Performances unidimensionnelles usuelles.....	61
2.	Scaling et Mixed Assessor Model	63
3.	Mixed Assessor Model (MAM).....	65
4.	Tests de performances prenant le scaling en compte	66
5.	La table MAM-CAP	67
6.	Discussion	68
II.	Les performances multidimensionnelles.....	69
1.	Le MultiMAM	69
2.	Tests de performances multidimensionnels	71
3.	Utilisation de la multi-dimensionnalité pour déterminer les causes du scaling	73
4.	Discussion	73
III.	Bilan des apports sur les performances du panel et des sujets	74
B.	Des cartes adaptées aux modèles statistiques proposés	76
I.	La CVA	76
1.	Concept	76
2.	Construction d'une CVA	77
3.	Exemple sur le jeu de données cheeses.....	78
4.	Discussion	79
II.	La MAM-CVA.....	80
1.	Principe.....	80

2.	Discussion	80
III.	A propos des ellipses de confiance.....	81
1.	Introduction.....	81
2.	Une nouvelle méthode pour construire des ellipses de confiance	82
3.	Exemple	83
4.	Discussion	83
IV.	Bilan sur les apports sur la cartographie des produits.....	84
C.	Conclusion du chapitre 3	85
Chapitre 4 :	Méta-analyses de la SensoBase	87
A.	Comparaison de cartes	89
I.	Méthodologie	89
1.	Un critère de comparaison des configurations des produits.....	89
2.	Deux critères de comparaison des interprétations sensorielles.....	89
3.	Un critère de comparaison des intersections des ellipses de confiance	90
4.	Un critère de comparaison du pouvoir discriminant multidimensionnel.....	91
II.	ACP contre CVA.....	91
III.	ACP contre CVA-MAM	92
IV.	Application de la méthodologie de comparaison de cartes à l'ACP verticale	93
V.	Discussion et bilan sur les comparaisons de cartes	93
B.	Performances.....	95
I.	Performances unidimensionnelles	95
1.	Performances du panel	95
2.	Performances unidimensionnelles du sujet.....	96
II.	Performances multidimensionnelles.....	97
1.	Paradoxe de la MANOVA.....	97
2.	Scaling.....	98
C.	Questions générales d'analyse sensorielle	100
I.	Intérêt des répétitions en analyse sensorielle	100
II.	A combien de sujets supplémentaires équivaut une répétition supplémentaire ?...	101
D.	Conclusion du chapitre 4	102
Chapitre 5 :	Synthèse et discussion.....	103
A.	Rappel de la démarche suivie	105
I.	Amélioration des calculs de performances	105

II. Amélioration des techniques de cartographies des produits	108
III. Utilisation des méthodologies mises en place pour des résultats d'analyse sensorielle	108
IV. Originalité de la démarche	108
B. Synthèse et discussion des résultats obtenus	110
I. Résultats méthodologiques	110
1. Méthodologies de méta-analyses	110
2. Méthodologie de comparaison de cartes	111
3. Méthodologie de comparaison de deux jeux de données de profil sensoriel	112
II. Résultats statistiques (et informatiques)	113
1. Tests de performances unidimensionnels et table MAM-CAP	113
2. Carte adaptée à l'analyse sensorielle : la CVA	114
3. Un modèle multidimensionnel prenant le scaling en compte	115
III. Synthèse et discussion sur les résultats d'analyse sensorielle	115
1. Pratiques d'analyse sensorielle mises en évidence par la Sensobase	115
2. Performances en profil sensoriel	116
3. Cartes en profil sensoriel	117
IV. Discussion générale, limites et perspectives	117
C. Recommandations à l'issue de ce travail	120
I. Recommandations sur le protocole de profil sensoriel	120
II. Recommandations sur l'analyse des données de profil sensoriel	120
1. Recommandations pour l'analyse des performances en profil sensoriel	120
2. Recommandations pour la construction de cartes produits en profil sensoriel	120
3. Synthèse des recommandations à destination de l'animateur de panel	121
Conclusion	123
Bibliographie	125
Table des figures et tableaux	129
Annexes	131

ANNEXES

Annexe 1 : Articles	133
A. The MAM-CAP Table : a new tool for monitoring panel performances.....	135
B. PCA versus CVA : Comparaison of 422 datasets.....	139
C. Canonical Variate Analysis of Sensory Data.....	147
D. Multiplicative decomposition of the scaling effect in the Mixed Assessor Model.....	171
E. Taking individual differences into account in the use of the scale with MultiMAM.....	187
F. Do we need to replicate in sensory profiling studies.....	207
Annexe 2 : Rapports de méta-analyses	221
A. Rapport R1 : ACP contre CVA, méta-analyse des biplots.....	223
B. Rapport R2 : Comparaison de la MAM-CVA et de l'ACP sur 379 jeux de données de la SensoBase	226
C. Rapport R3 : Comparaison de l'ACP et de l'ACP verticale par méta-analyse de la SensoBase	229
D. Rapport R4 : Etude des performances du panel sur 177 jeux de données.....	234
E. Rapport R5 : Performances des sujets en fonction de leur caractéristiques.....	239
Performances des panelistes en fonction du genre	242
performances des sujets en fonction du statut tabagique.....	243
performances des sujets en fonction de l'âge.....	244
F. Rapport R6 : Equivalence entre nombre de sujets supplémentaires et ajout d'une répétition	245
Annexe 3 : Aide des packages TimeSens	255
A. Aide du package CVA.....	256
B. Aide du package MAMCAP	261

PREAMBULE

En 1956, le tout premier disque dur était créé avec une capacité de stockage de 5 Mo. Près de 60 ans plus tard, cette capacité a été multipliée par plus d'un million. En parallèle, de nouveaux outils de collecte de données (internet, téléphones portables, ordinateurs de plus en plus puissants...) sont apparus. Ces progrès fulgurants permettent aujourd'hui à la plus petite structure de collecter et de stocker des quantités de données impressionnantes. Ces données, qu'elles soient structurées (**bases de données**) ou non, constituent une richesse bien souvent inexploitée faute d'outils méthodologiques et d'expertises statistiques.

L'analyse sensorielle n'échappe pas au phénomène. Cette discipline permet de mesurer la perception sensorielle de produits, en général alimentaires, en utilisant l'être humain comme outil de mesure. Pour cela, un ensemble de personnes (appelées indifféremment juges, sujets ou panélistes) est convoqué par un « animateur de panel » qui lui fait tester des produits. L'épreuve de profil sensoriel est un protocole consistant à attribuer des notes d'intensité de différents descripteurs (sucré, brillant, mou...) pour chaque produit une ou plusieurs fois selon un ordre défini par un plan d'expérience.

Une fois les données collectées, l'animateur de panel doit les analyser statistiquement pour aboutir à une caractérisation sensorielle de ses produits. Pour choisir quelle méthode statistique utiliser, l'animateur de panel peut se baser sur différents éléments théoriques (recommandations des normes d'analyse sensorielle, bibliographie) et/ou pratiques (coût éventuel du logiciel à utiliser, disponibilité de la méthode, facilité d'utilisation de cette méthode). L'analyse statistique finalement effectuée n'est donc pas forcément la plus pertinente théoriquement, mais souvent celle qui est la plus courante et la plus accessible.

Les objectifs de cette thèse sont de déterminer quelles sont les méthodes les plus pertinentes pour l'analyse de profil sensoriel, puis de les comparer aux méthodes les plus couramment utilisées. Ce travail sera effectué à l'aide de l'exploitation d'une base contenant plus de mille jeux de données de profil sensoriel (on parlera de **méta-analyse**). Cette approche permettra également de quantifier l'impact de certaines caractéristiques de l'étude sur les résultats observés : âge et genre des panélistes, modalité sensorielle des descripteurs (visuelle, gustative,...). Des recommandations à destination des animateurs de panel seront ainsi établies.

Les méthodes statistiques développées dans cette thèse seront valorisées par leur implémentation dans des bibliothèques R mises à disposition des animateurs de panel. Ces bibliothèques seront également utilisables via le logiciel TimeSens à destination des animateurs de panels ne maîtrisant pas le logiciel R.

Pour que ce travail de thèse reste compréhensible par tout animateur de panel (même non statisticien), les éléments théoriques de statistiques et les détails des méta-analyses seront placés dans des paragraphes spécifiques (généralement en annexe) et leur lecture ne sera pas nécessaire à la compréhension globale du manuscrit.

CHAPITRE 1 :

INTRODUCTION

A. ANALYSE SENSORIELLE

I. PRESENTATION DE L'ANALYSE SENSORIELLE

La norme française NF ISO 5492 définit l'analyse sensorielle comme « l'examen des propriétés organoleptiques d'un produit par les organes des sens ». L'analyse sensorielle est utilisée par les industriels agro-alimentaires principalement, mais aussi par les industriels dans le domaine des cosmétiques et de l'automobile. Ses objectifs peuvent être divers (conception produit, démarche qualité, positionnement d'un produit par rapport à la concurrence...) mais la démarche est commune. Ainsi, lors d'une épreuve d'analyse sensorielle, les **panélistes** (ou **sujets**) goûtent ou testent différents **produits** et les évaluent à l'aide d'un ou plusieurs **descripteurs** (caractéristiques des produits) une ou plusieurs fois (**répétitions**).

Il s'agit donc d'utiliser l'être humain comme **instrument de mesure** en mettant à profit ses capacités olfactives, gustatives, visuelles, auditives et tactiles pour **caractériser et évaluer des produits**.

II. L'ETRE HUMAIN COMME OUTIL DE MESURE

1. L'ETRE HUMAIN, UN INSTRUMENT DE MESURE DES CINQ SENS

Lorsqu'il déguste un aliment, le panéliste est un instrument de mesure idéalement capable de détecter :

- la **saveur** de l'aliment par le **goût** (sucré, salé...),
- l'**aspect visuel** de l'aliment par la **vue** (lisse, jaune...),
- la **texture** de l'aliment par le **toucher** (mou, dur, fondant...),
- les **arômes** libérés par certaines molécules volatiles odorantes par l'**odorat** (cacao, fumé...),
- des **bruits** de mastication par l'**ouïe** (croustillant, craquant, pétillant...),
- certains stimuli **trigéminaux** ou chimiques difficiles à séparer des sensations gustatives ou olfactives mais causant des perceptions de brûlure, de chaleur, de froid ou de piquant dans les muqueuses du nez et de la bouche.

Ainsi, tous les sens sont mobilisés pendant la dégustation d'un aliment. Les phénomènes biologiques impliqués dans la perception sensorielle sont connus et se décomposent en trois étapes : la **stimulation** par les récepteurs sensoriels concernés, la **transduction** du message chimique en message nerveux et l'**interprétation** du message nerveux par le cerveau qui utilise l'expérience qu'il a en mémoire pour convertir la sensation en perception.

Le **goût**, dans son sens strict, correspond à l'activation de récepteurs situés sur la langue par des molécules sapides : des bourgeons situés sur les papilles contiennent des récepteurs permettant de discerner les 5 saveurs fondamentales : le salé, le sucré, l'amertume, l'acidité

et l'umami. Puis, les récepteurs transforment le signal chimique en un signal nerveux conduit dans la zone corticale du cerveau (Figure 1).

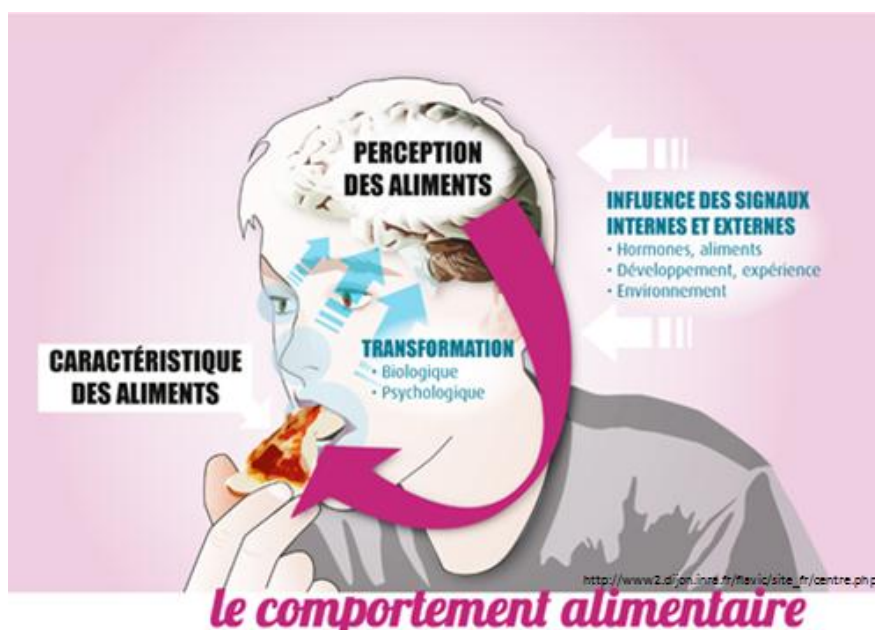


FIGURE 1: MECANISMES DU GOUT

En ce qui concerne la **vue**, la lumière (onde électromagnétique) entre dans l'œil et va activer des récepteurs dans la rétine, appelés « photo récepteurs ». Ces récepteurs transforment le signal électromagnétique en signal électrique (ou message nerveux) (Meilgaard, Gail Vance Civile, & Carr, 1999). Ce signal est ensuite transmis par les neurones au système nerveux central par le nerf optique qui conduit le signal vers la zone visuelle du cortex cérébral.

Dans le cas de l'**odorat**, des molécules olfactives stimulent des récepteurs olfactifs dans la cavité nasale (épithélium olfactif). Chaque type de récepteur olfactif (il en existe plus de mille) possède une sensibilité particulière. Puis, les neurones olfactifs dirigent le message vers le bulbe olfactif, dans la région préfrontale du cerveau. La manière dont les récepteurs génèrent les signaux qu'ils envoient au cerveau reste mal connue. Lors de la dégustation d'un aliment, on distingue l'odorat orthonasal (intervenant lorsque l'on sent l'aliment) de l'odorat rétronasal (intervenant lorsque l'aliment est en bouche, et que les molécules odorantes suivent la voie rétro-nasale pour rejoindre la cavité).

Le **toucher** est basé sur la présence de récepteurs situés sous la peau, répondant à différents types de stimulus (chaleur, froid, pression ou douleur). Des terminaisons nerveuses situées dans l'épiderme transforment les informations obtenues par les récepteurs sensoriels en messages nerveux, qui sont ensuite conduits par la moelle épinière jusqu'au cerveau.

L'**ouïe** interprète un stimulus basé sur la vibration de l'air : cette vibration est amplifiée et véhiculée jusqu'à l'oreille interne, qui contient un grand nombre de cellules réceptrices munie de cils microscopiques. Les ondes mécaniques font bouger les cils, activant la

production de messages nerveux. Ces messages nerveux sont ensuite conduits par le nerf auditif jusqu'au cortex auditif du cerveau.

En plus d'être le consommateur final du produit, l'être humain est donc un instrument de mesure complexe qui présente l'avantage d'être peu coûteux et disponible comparé à d'autres outils pouvant potentiellement décrire un aliment (nez ou bouche artificiels, analyses physico-chimiques...). Il est capable de capter un très grand nombre d'informations sur l'aliment, dont la plupart sont difficilement mesurables avec des instruments de mesure classiques (odeur de fraise, sensation de gras...). Cependant, l'étude des perceptions de l'être humain ne permet pas d'avoir un diagnostic sur la composition physico-chimique d'un produit (et réciproquement).

2. L'ÊTRE HUMAIN : UN INSTRUMENT DE MESURE SOUMIS A DIFFERENTS BIAIS

Cependant, et c'est tout le problème en analyse sensorielle, **les réponses de deux individus au même stimulus peuvent être (et sont la plupart du temps) différentes**. Plusieurs raisons sont évoquées dans Meilgaard et al. (1999) :

- une **différence de sensibilité** (due à l'organe sensoriel propre à chaque individu),
- une **différence dans le traitement de cette sensation** (par exemple à cause d'un manque de connaissances sur une odeur particulière),
- un **niveau d'entraînement différent dans l'expression de la sensation** en chiffres et en mots. A travers l'entraînement et l'usage de références, l'animateur de panel espère modeler le procédé mental de manière à ce que les sujets donnent à peu près la même réponse à un stimulus donné.

Lim (2011) a proposé le schéma suivant (Figure 2) afin de représenter les différentes étapes de l'obtention d'une note d'un stimulus sensoriel.

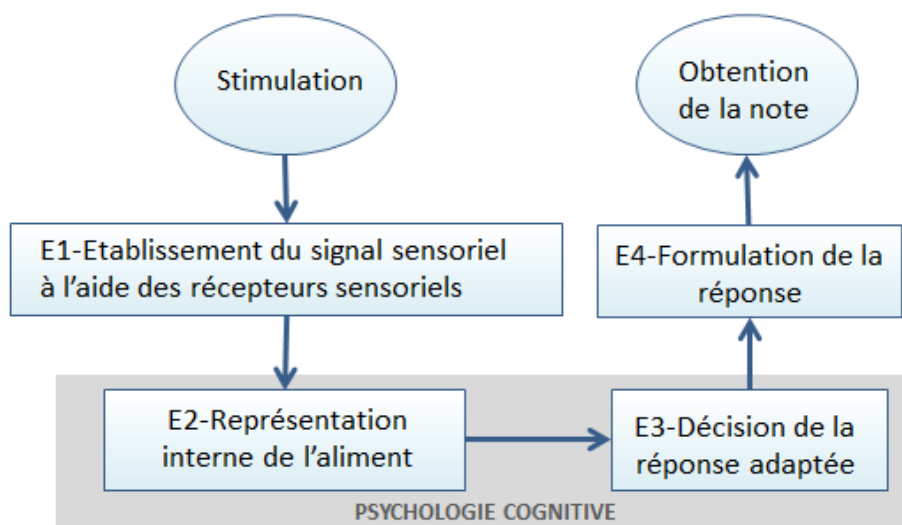


FIGURE 2 : ETAPES DE L'OBTENTION D'UNE NOTE APRES STIMULATION SENSORIELLE

Pour résumer, les notes obtenues en analyse sensorielle peuvent être soumises à **différents biais psychologiques** – biais dû à la représentation interne de l'aliment (E2) ou à la décision de la réponse adaptée (E3) – **ou physiologiques** – biais dus aux récepteurs sensoriels lors de la dégustation (E1) ou à la capacité physique de restituer la note choisie sur une échelle (E4) –. Nous verrons dans la suite de cette thèse qu'il est possible de corriger certains de ces biais dans le traitement statistique des données d'analyse sensorielle.

Une particularité de l'analyse sensorielle est qu'un même sujet n'attribuera pas forcément la même note à un même stimulus présenté plusieurs fois : les biais décrits ci-dessus ne sont pas uniquement valables pour décrire des différences inter-individuelles, mais aussi des différences intra-individuelles. Cette réalité doit absolument être prise en compte en analyse sensorielle.

i. BIAIS PSYCHOLOGIQUES

Stone and Sidel (2004) et Meilgaard et al. (1999) répertorient différents types de biais psychologiques pouvant avoir lieu lors d'évaluations sensorielles. On peut classer ces principaux biais de la manière suivante :

- Les **biais de supposition** : si les sujets ont des attentes sur ce qu'ils doivent percevoir lors de leur dégustation pour différentes raisons (connaissance des produits, entraînement avec descripteurs suggestifs, apparence des produits trompeuse...).
- Les **biais d'habitude** : lorsqu'un panéliste a tendance à répéter les mêmes scores (par exemple dans un contrôle qualité quotidien).
- Les **biais induits par l'ordre de présentation des échantillons** : effet lié à l'ordre des produits (effet de contrastes) ou lié au moment de leur présentation (le premier produit est souvent préféré par rapport aux autres) ou encore au fait qu'on place les premiers produits au centre de l'échelle pour avoir de la place pour répondre par la suite (biais de tendance centrale).
- Les **biais dus au manque de motivation** : un panéliste intéressé aurait des réponses plus pertinentes qu'un panéliste non motivé, qui pourrait prendre l'expérimentation à la légère et ne pas répondre sérieusement aux questions.
- Un **biais de dispersion des notes** : certains panélistes ont tendance à davantage sous- (ou sur) disperser leur notes que les autres, quelle que soit l'échelle fixée.

D'autres biais, telles que le biais de suggestion mutuelle lorsque les sujets ne notent pas les produits individuellement, ou l'effet de halo (influences des notes attribuées entre elles) sont également cités.

L'animateur de panel doit veiller à réduire au maximum ces risques de biais et pour cela, il a différentes réponses possibles résumées dans le Tableau 1.

Source de biais	Réponse de l'animateur de panel
Biais de supposition	Maitriser les informations transmises au sujet
Biais d'habitude	Eviter les tâches répétitives, varier les tests
Ordre de présentation	Présenter les produits suivant un ordre randomisé et équilibré
Manque de motivation	Insister sur l'importance de la qualité de travail des panélistes, rémunérer les juges
Biais de dispersion des notes	Homogénéiser l'utilisation de l'échelle à l'aide de références

TABLEAU 1 : REPONSES POSSIBLES DE L'ANIMATEUR DE PANEL A DIFFERENTES SOURCES D'ERREURS D'ORIGINE PSYCHOLOGIQUES

ii. BIAIS PHYSIOLOGIQUES

L'**adaptation** constitue un premier biais physiologique. Elle consiste en une diminution ou un changement de sensibilité à un stimulus donné dû à une exposition répétée ou continue à ce stimulus. Elle peut être corrigée en limitant le nombre de produits présentés à la suite lors d'une séance d'analyse sensorielle, ou en proposant des verres d'eau entre les différents produits.

L'**interaction de plusieurs stimuli** peut également être à l'origine d'un biais. Ainsi l'effet de la présence d'une substance peut augmenter (synergie) ou diminuer l'intensité d'une seconde substance.

iii. BIAIS DUS A L'ETAT DE SANTE DU PANELISTE

Les panélistes ne participent (théoriquement) pas aux études s'ils souffrent de fièvre, de rhume, de problèmes dentaires ou encore de bouleversements émotionnels qui les empêchent de se concentrer.

iv. CARACTERISTIQUES DU SUJET ET APTITUDES SENSORIELLES

Certaines caractéristiques telles que le fait de fumer ou l'âge des panélistes peuvent également jouer un rôle dans leurs aptitudes sensorielles.

En ce qui concerne le fait de fumer, Krut (1961) a comparé les seuils de détections des saveurs salées, sucrées, acides et amères, pour un groupe de fumeurs et un autre de non-fumeurs. Des différences significatives ($p < 0.001$) ont été déterminées uniquement pour les saveurs amères. Chez les fumeurs, on a observé une détérioration de la détection de l'amertume, et les petits fumeurs (moins de 10 cigarettes par jour) semblent moins affectés que les gros fumeurs (plus de 10 cigarettes par jour). Cette étude a également montré que fumer une cigarette ne semblait pas avoir d'effet immédiat sur la perception du goût. Malgré cela, lors de la plupart des études sensorielles, il est recommandé de ne pas fumer 30 minutes avant une dégustation. Par ailleurs, Ahlström, Berglund, Berglund, Engen, and Lindvall (1987) ont étudié dans quelle mesure la perception des intensités des odeurs variait pour les fumeurs et les non-fumeurs. Il a été montré que les fumeurs étaient moins sensibles aux odeurs que les non-fumeurs. Les auteurs proposent deux explications à cette hypo-

sensibilité : une explication psychologique (les fumeurs attribuent le terme d'odeur à des perceptions plus intenses), ou bien, et c'est leur explication privilégiée, une perte de capacité physiologique à détecter les odeurs.

Booth, Kohrs, and Kamath (1982) ont présenté une revue bibliographique des liens entre acuité (ou sensibilité) gustative et âge. Il a été montré que le nombre de bourgeons gustatifs et la sensibilité aux saveurs sucrée et salée semblent diminuer avec l'âge, et que la sensibilité à l'amertume était corrélée au fait de fumer ou non.

III. MISE EN PLACE D'UNE ETUDE SENSORIELLE

Pratiquement, une étude sensorielle doit être menée en suivant différentes étapes (Enhardt, 1978) :

- définir l'objectif du test,
- choisir les échantillons,
- élaborer l'expérimentation,
- mener à bien l'expérimentation,
- analyser les données obtenues,
- interpréter les résultats des analyses.

Le type d'épreuve sensorielle est déterminé en fonction de l'objectif principal de l'étude (Tableau 2). Il existe également des expérimentations mixtes, comprenant différents types d'épreuves. Par exemple, on peut chercher à expliquer les préférences par des caractéristiques sensorielles.

Dans cette thèse, nous nous intéresserons exclusivement aux épreuves descriptives.

Type d'épreuve sensorielle	Objectif principal de l'étude
Epreuve de discrimination	Déterminer si des produits sont perçus identiques ou non
Epreuve descriptive	Déterminer les caractéristiques sensorielles des produits
Epreuve hédonique	Déterminer les niveaux de préférences des produits d'une cible de consommateurs

TABLEAU 2 : TYPE D'EPREUVES SENSORIELLES EN FONCTION DE L'OBJECTIF DE L'ETUDE

IV. TIMESENS, UN LOGICIEL D'ACQUISITION ET D'ANALYSE DES DONNEES D'ANALYSE SENSORIELLE

La plupart des animateurs de panel ont informatisé l'acquisition et le traitement des données d'analyse sensorielle à l'aide d'un des logiciels dédiés à l'analyse sensorielle : FIZZ, EyeQuestion, CompuSens...

Développé par la plateforme ChemoSens du Centre des Sciences du Goût et de l'Alimentation (CSGA), le logiciel **TimeSens** permet l'**acquisition** et le **traitement** de données sensorielles. Il a été d'abord conçu pour l'exploitation des données de Dominance Temporelle des Sensations (DTS), une épreuve descriptive prenant en compte la temporalité des sensations (Pineau & Schlich, 2015), puis a été étendu à d'autres protocoles, dont le profil sensoriel.

Les analyses statistiques sont effectuées à l'aide d'un service Web qui utilise une **librairie R** (logiciel libre de traitement de données et d'analyse statistique, <http://cran.r-project.org/>) spécialement développée, et dans laquelle il est facile d'intégrer des nouveaux traitements statistiques.

Etant donné que la plupart des animateurs de panel choisissent souvent leurs analyses **en fonction de leur disponibilité et de leur facilité d'utilisation** (et pas forcément en fonction de leur pertinence par rapport à leurs données), tous les modèles proposés et discutés dans la thèse seront implémentés dans la librairie du logiciel TimeSens et mis à disposition une fois publiés. Les animateurs de panel qui n'utilisent pas R pourront accéder à ces analyses via TimeSens avec une interface facile à prendre en main.

B. ANALYSE DESCRIPTIVE QUANTITATIVE (QDA)

I. DEFINITION ET HISTORIQUE DES ANALYSES DESCRIPTIVES

1. DEFINITION

L'analyse descriptive consiste à faire **évaluer les différentes caractéristiques sensorielles d'un produit par un panel de sujets entraînés**. Son objectif est de mettre en place des expérimentations permettant d'aboutir à une description la plus complète possible des produits (ISO 11035) :

- sur le plan qualitatif, en définissant à l'aide de descripteurs l'ensemble des perceptions permettant de discriminer un produit par rapport à d'autres du même type,
- sur le plan quantitatif, en évaluant l'intensité de chaque descripteur.

La méthode du profil sensoriel peut s'appliquer (ISO 11035) pour définir un standard de fabrication, améliorer/développer des produits, visualiser l'influence du vieillissement des produits et des conditions de stockage ou encore comparer un produit avec ceux du même type déjà commercialisés.

2. AVANT LE PROTOCOLE QDA : HISTORIQUE DE L'ANALYSE DESCRIPTIVE

La méthode « **Flavour Profile** », présentée par Cairncross and Sjöström (1950), fut la première procédure descriptive qualitative de test sensoriel, introduite dans les années 1950. Cette méthode utilise un panel de 4 à 6 sujets sélectionnés basé sur des séries de tests d'acuité sensorielle, leur intérêt pour l'analyse sensorielle et leur disponibilité. Les sujets examinent et discutent le produit en session ouverte, et une fois qu'un accord concernant la description du produit est atteint, l'animateur de panel résume les résultats sous forme d'un rapport (Stone & Sidel, 2004). Plus tard, la méthode « **Texture Profile** » (Brandt, Skinner, & Coleman, 1963) introduisit dans le Flavour Profile Method des échelles de notation en 13 points (puis continues) pour chaque terme de texture.

L'utilisation des échelles fut ensuite étendue à l'ensemble des descripteurs dans la méthode **Quantitative Descriptive Analysis (QDA)** (Stone, Sidel, Olivier, Woolsey, & Singleton, 1974).

3. LA METHODE QDA

Les panélistes **évaluent les produits un par un** (profil monadique) dans des cabines séparées pour réduire la distraction et l'interaction potentielle entre panélistes. Ils notent la série de produits sur différents descripteurs sensoriels. Notons également qu'une épreuve de profil comparative existe (dans ce cas, tous les produits sont présentés simultanément, et chaque descripteur est noté un par un).

Cette méthode est caractérisée par les points suivants (Urdapilleta, Ton Nu, Saint Denis, & Huon de Kermadec, 2001) :

- Les panélistes sont sélectionnés parmi un grand nombre de candidats en fonction de leur capacité à discriminer les différences dans les propriétés sensorielles des échantillons d'un type de **produit spécifique** pour lesquels ils sont **entraînés**.
- L'entraînement d'un panel QDA nécessite **l'utilisation de références** et le **développement d'une terminologie commune**. La mise en place de la liste de descripteurs résulte d'un **travail de groupe**.
- Les évaluations sont recueillies (en version papier) sur une **échelle continue linéaire** (15 cm de long avec encoche à 1.3 cm de chaque extrémité et au centre).
- Les **performances des sujets** sont vérifiées au cours des différentes phases de l'étude.

Notons qu'en pratique, certains de ces points ne sont pas toujours mis en place mais que cette méthode est aujourd'hui très largement répandue.

4. AUTRES PROTOCOLES ISSUS DE LA METHODE QDA

La méthode QDA a été adaptée par la suite (à d'autres échelles de notation et d'autres modes de présentation des échantillons...) jusqu'à aboutir au **profil conventionnel**, la méthode recommandée par les normes AFNOR (AFNOR, ISO 13 299, juin 1995) (Urdapilleta et al., 2001).

La **méthode « Spectrum »** (Muñoz & Civille, 1992) permet aux panélistes de noter leurs perceptions sur une **échelle « universelle »**. Cette échelle est basée sur l'utilisation de références existant pour différents types de descripteurs. Son avantage principal est qu'elle permet de comparer un produit avec un autre produit testé par un autre laboratoire à une autre période.

Un **grand nombre d'alternatives au profil sensoriel QDA** ont été développées pour remplacer le panel (coûteux à entraîner) par des consommateurs : Free Choice Profiling (Williams & Arnold, 1984), Flash Profiling (Delarue, 2015), Check All That Apply (CATA) (Ares & Jaeger, 2015), Ideal Profiling (Worch & Punter, 2015), Projective Flash Profile (Ballay, Loescher, & Gazano, 2015). Ces méthodes alternatives dites « rapides » concurrencent directement le profil sensoriel classique. En effet, même si l'efficacité de ces méthodes n'est pas toujours reconnue, elles sont plus rapides et moins coûteuses.

Des **méthodes temporelles** ont également été proposées : la méthode du Temps Intensité (TI) (Larson-Powers & Pangorn, 1978) ou la méthode de Dominance Temporelle des Sensations (DTS) (Pineau, Cordelle, & Schlich, 2003; Pineau et al., 2012).

Aujourd'hui, dans la pratique, les méthodes QDA et profil conventionnel sont souvent confondues. Nous les regrouperons ici sous le terme **profil sensoriel**. Cette thèse ne s'intéressera qu'aux protocoles et analyses de profil sensoriel.

II. QUESTIONS CLASSIQUES SUR LE PROTOCOLE DE PROFIL SENSORIEL

1. ENTRAÎNEMENT DES SUJETS

Lors de leurs études de profil sensoriel, les animateurs de panels utilisent un **panel d'experts**, entraînés à évaluer une gamme de produits spécifique. D'autres entraînent un **panel *a priori* naïf**. Les descripteurs sont définis soit par l'animateur de panel, soit par le panel durant l'entraînement. Une phase de génération de descripteurs par le panel est souvent la première étape de l'entraînement, comme conseillé dans Stone and Sidel (2004). Puis, on procède théoriquement à une phase de vérification pour établir que les sujets sont bien capables de reconnaître les descripteurs, puis de discriminer les produits (c'est-à-dire établir des différences entre des échantillons différents). Certains animateurs de panel conservent uniquement les sujets discriminants pour effectuer leur études, d'autres ne font pas de sélection et les phases d'entraînement sont plus ou moins longues en fonction de la complexité des produits évalués.

2. NOMBRE DE SUJETS, PRODUITS, DESCRIPTEURS, REPETITIONS

Lors d'un profil QDA, une **note** est généralement attribuée pour chaque 5-uplet **panéliste x produit x descripteur x répétition x session**. Une session correspond en général à la date de l'expérimentation alors que la répétition se rapporte au nombre de fois où le produit est dégusté. Théoriquement, une répétition devrait systématiquement être interne à une même session. En d'autres termes, l'ensemble des répétitions des produits devraient être effectuées lors de la même séance. Dans la pratique, il est courant que les répétitions aient lieu lors de sessions différentes à cause de contraintes pratiques (nombre de produits trop élevé pour les évaluer plusieurs fois en une seule séance). Par conséquent, dans cette thèse, on considèrera plusieurs répétitions lorsqu'un même produit est noté plusieurs fois pour un même descripteur et un même sujet (peu importe la session). Cela signifie que **l'effet de la session sera négligé dans nos travaux**.

Le nombre de répétitions a été discuté par Mammasse (2012), qui affirme que, si les répétitions sont utiles en phase d'entraînement, elles ne le sont plus forcément autant lors de la période de mesure. Concernant le nombre de sujets à enrôler dans un panel de profil sensoriel, il existe différentes recommandations qui convergent vers un panel de 10 à 12 sujets. Ces chiffres ont été validés par méta-analyses (Mammasse, 2012).

Le nombre d'échantillons à faire déguster au panéliste dépend fortement de la catégorie de produit, de la quantité à ingérer et de la fatigue/saturation sensorielle engendrée par l'ingestion du produit.

Le nombre de descripteurs est également très variable. Ces descripteurs peuvent être regroupés par « modalité sensorielle » en fonction des sens mobilisés (liste non exhaustive) :

- les descripteurs visuels (aspect, couleur...),
- les descripteurs de texture (friable, dur...),
- les descripteurs de saveur (sucré, salé...),
- les descripteurs d'arôme (fruité, chocolat...),
- les descripteurs d'odorat (odeur de fraise, d'herbe coupée...),
- les descripteurs trigéminaux (piquant, mentholé, brûlant ...),
- les descripteurs d'arrière-goût.

III. ANALYSES CLASSIQUES DES DONNEES DE PROFIL SENSORIEL

Généralement, les données QDA sont analysées pour répondre à différentes questions (Tableau 3).

Question d'analyse sensorielle	Analyse associée	Commentaires
Mon panel et mes panélistes sont-ils performants?	variable	Ne prend pas en compte l'hétérogénéité de la dispersion des notes des panélistes
Comment représenter graphiquement les données de l'expérience ?	ACP	Ne prend pas en compte la variabilité des sujets autour de la moyenne
Les produits sont-ils différents ?	ANOVA	Ne prend pas en compte l'hétérogénéité de la dispersion des notes des panélistes
Existe-t-il des groupes de produits?	Test HSD	Ne prend pas en compte l'hétérogénéité de la dispersion des notes des panélistes

TABLEAU 3 : TRAITEMENTS STATISTIQUES EN ANALYSE SENSORIELLE

Certaines de ces questions sont posées pour chaque descripteur (traitement unidimensionnel des données) ou tous descripteurs confondus (traitement multidimensionnel des données). La plupart de ces questions se posent pour les deux approches. Lors du traitement multidimensionnel des données, il peut être pertinent d'analyser séparément les modalités sensorielles (et d'obtenir ainsi des résultats indépendants sur l'aspect visuel, le goût, les odeurs...).

1. PERFORMANCES

La notion de performance en analyse sensorielle est le plus souvent décomposée en trois concepts : la discrimination, la répétabilité et l'accord entre les juges (Pineau, 2006). On distingue les performances du panel de celles des juges.

La **répétabilité** traduit l'aptitude à donner des notes très proches à un même stimulus. Elle est appelée reproductibilité lorsque les conditions de présentation du même stimulus sont

différentes. Le contrôle de la répétabilité ou de la reproductibilité du panel est indispensable pour attester de la qualité des résultats obtenus et/ou à obtenir.

La **discrimination** est la capacité à détecter des différences entre des produits différents, cette performance est donc au centre des questionnements de l'animateur de panel sensoriel.

L'**accord** entre les juges permet de vérifier que les panélistes sont parvenus à un consensus et qu'ils sont globalement d'accord entre eux sur les notes à attribuer aux produits.

L'étude de la bibliographie n'a pas établi de méthode générale pour les tests de performances des individus. La discrimination individuelle est souvent testée pendant l'entraînement et permet de déterminer si le sujet participera aux séances de mesure.

Schlich (1997a) a présenté une table **résumant l'ensemble des performances du panel et des panélistes** de manière synthétique avec des codes couleur : la **table CAP** (Control of Assessor Performances). Elle a depuis été utilisée par de nombreux industriels grâce à un projet d'échange de jeux de données de profil sensoriel contre des analyses statistiques (www.sensobase.fr) (Pineau, 2006).

2. CARTES SENSORIELLES

Une carte sensorielle permet de visualiser rapidement l'ensemble des résultats d'une expérimentation de profil, tous descripteurs et produits confondus. Cette représentation graphique est par conséquent très prisée par les animateurs de panel. La technique de cartographie la plus souvent utilisée est l'**Analyse en Composantes Principales (ACP)** du tableau des moyennes des produits en lignes et des descripteurs en colonnes (cf. Chapitre 1. C. 2).

3. DIFFERENCES ENTRE PRODUITS

Pour visualiser la distribution des notes, différentes méthodes peuvent être utilisées (histogramme, boîtes à moustaches, tableau d'information, radar plot...). Cependant, elles ne comportent aucun test et ne permettent pas de conclure statistiquement à des différences entre produits. L'**Analyse de la Variance (ANOVA)** (cf. Chapitre 1. C. 3) va permettre de déterminer si les moyennes de tous les produits sont statistiquement différentes pour un descripteur fixé (dans ce cas on parlera d'un effet produit significatif). Une fois qu'un effet produit a été mis en évidence, un test de différence *a posteriori* comme le test de Tukey (Honest Significance Differences Test ou HSD test) est généralement conduit pour identifier les groupes de produits avec des moyennes statistiquement différentes.

IV. REVUE BIBLIOGRAPHIQUE DES ETUDES QDA REALISEES EN 2014

Afin d'avoir un aperçu des pratiques courantes du profil QDA (protocole et analyses) en analyse sensorielle, nous avons recherché les réponses à une requête « QDA profile » sur sciencedirect, une plateforme de publications académiques gérée par l'éditeur Elsevier.

Parmi 367 articles détectés, 38 réponses correspondant réellement au profil QDA ont été obtenues en 2014. Les caractéristiques de ces études ont été listées dans le tableau. Quatre objectifs principaux ont été répertoriés pour les études de profil sensoriel :

- L'impact d'un changement (conditions de stockage, de cuisson, de préparation) sur le produit (P)
- Une mise en relation avec les propriétés physico-chimiques des produits (C)
- Une mise en relation entre les propriétés sensorielles et les préférences (L)
- Une question méthodologique (M)

Les caractéristiques des 38 jeux de données ont été répertoriées dans le Tableau 4. Ces jeux de données proviennent principalement de Food Control (FC) (9/38), Food Science and Technology (FST) (7/38), Meat Science (MS) (5/38), Food Research International (FRI) (4/38), PostHarvest Biology and Technology (PBT) (3/38) et Food Microbiology(FMI) (3/38). Sept autres journaux ont proposé uniquement un article sur le profil QDA en 2014 : Food Hydrocolloids (FH), Journal of Dairy Science (JDS), Food Chemistry (FCh), Food Bioscience (FB), Journal of Bioscience and Bioengineering (JBB), Small Ruminant Research (SRR) and International Dairy Journal (IDJ) .

Les caractéristiques des 38 études ont été répertoriés dans le Tableau 4.

1 ^{er} auteur	Journal	Type de produit	Obj.	NS	NP	NR	NSe	NA	ACP	ANOVA	HSD
Biasoto	FRI	Vins rouges	C,L	12	10	6	5	33	X	X	X
Lorido	MS	Viande ibérienne	C,M	11	3	3	3	31			
Egea	MS	Viande	P	8	18	1	3	7			X
Devezeaux de Lavergne	FH	Gels	M	12	8	3	2	36	X	X	
Morais	FST	Pains sans gluten	L	15	6	4	?	15		X	X
Pimentel	FST	Jus de pommes sans probiotiques	L	18	7	3	?	11	X	X	
Amoa-Awua	FC	Dawadawa (Condiment d'Afrique de l'Ouest)	P	14	6	2	?	11	X		
Sun	FMI	Vins à la cerise	P	7	6	2	?	7	X	X	
Chojnicka-Paszun	FRI	Gel	C	11	18	2	6	5			
Francis	FST	Chicorée	P,C	8	6	2	4	26	X	X	X
Gomes	FRI	Steaks	P,L	13	6	6	6	23	X	X	X
Rizzolo	PBT	Poires	P	10	12	1	3	7		X	X
Feng	FC	Sauce soja	P	7	10	3	?	10		X	
Genisheva	FST	Vin	P	10	7	1	1	29			
Bagdi	FST	Pâtes	P	14	2	?	?	12			

Chapitre 1 : Introduction

1 ^{er} auteur	Journal	Type de produit	Obj.	NS	NP	NR	NSe	NA	ACP	ANOVA	HSD
Paulsen	MS	Saucisses	P	10	9	2	5	19		X	X
Picouet	FST	Jus de carotte	P	6	4	2	8	12		X	
Esposito	JDS	Caciocavallo fromage	P	12	3	5	10	17			
Castro-Vazquez	FC	Miel Lavande	P	8	26	2	?	13			
Rigou	FCh	Vin rouge	C	31	10	?	?	1		X	X
Rai	FB	Boissons	P,C	15	2	1	1	7		X	X
Cadena	FRI	Yaourts	M	9	8	2	?	13	X	X	X
Feng	FC	Sauce chinoises	C	8	2	?	7	7		X	X
Hixson	FC	Poissons	P	24	2	?	?	7	X		
Laguna	FST	Biscuits	P	8	4	2	?	17		X	X
Aquino	FRI	Lait	C	13	5	3	3	9		X	
Quero	FMi	Lait fermenté	C	6	4	2	1	13		X	
Lu	FST	Nouilles	P	10	?	?	?	7			
Yamamoto	JBB	Sauce soja	C	13	19	1	1	5			
Lorentzen	FC	Crabe	P,C	5	15	5	5	10	X		
Salvador	SRR	Fromage chèvre	P	11	2	2	2	10		X	X
CSarnacka-Szymani	IDJ	Fromage	P	6	3	?	?	18		X	
Canto	MS	Steak de caiman	P,C	8	5	1	1	11	X	X	
Silvano	FC	Miels	C	8	24	?	6	6			
Bechoff	FC	Boisson ibiscus	C	17	8	3	?	11	X	X	
Wojciak	MS	Moutarde/saucisse	P	7	4	2	?	15			X
Latocha	PBT	Kiwi	P	12	4-6	2	6	17	X	X	
Trzaskowska	FC	Saucisses	P	9	?	?	?	19	X		

TABLEAU 4 : CARACTERISTIQUES DES 38 ETUDES DE PROFIL SENSORIEL EFFECTUEES EN 2014

Avec NS le nombre de sujets, NP le nombre de produits, ND le nombre de descripteurs, NR le nombre de répétitions, NSe le nombre de sessions, Obj. leurs objectifs. Les trois dernières colonnes décrivent les analyses statistiques effectuées.

Cette étude bibliographique nous montre que la plupart des études publiées en 2014 ont pour but de mieux connaître les produits (24/38) ou de les mettre en relation avec leurs propriétés physico-chimiques (14/38).

Les caractéristiques des jeux de données semblent être assez variables suivant les études (nombre de sujets de 5 à 31, nombre de produits de 2 à 24, nombre de descripteurs de 5 à 36).

On remarque que seulement 11/38 études ont utilisé un panel de la taille recommandée (10 à 12 sujets) et 22/38 études avec au moins dix sujets.

Les analyses observées dans ces articles sont principalement des ANOVA (23/38). A la suite de ce test pour distinguer des différences, un test de Tukey (HSD) est le plus souvent réalisé (14/38) pour déterminer dans quel sens vont ces différences.

L'ACP permet de réaliser une cartographie de l'espace des produits pour 14/38 des cas. Lors de l'étude, d'autres analyses ont également été observées ponctuellement, telles que des tests de Student (t-tests), des radarplots ou d'autres analyses spécifiques au couplage analyse physico-chimique / analyse sensorielle ou profil / liking.

Cependant, il reste à préciser qu'il ne s'agit ici que des études publiées et à vocation « recherche ». Cette catégorie d'étude n'est pas forcément représentative de ce qui est fait en industrie, mais on suppose que cela reste quand même un bon indicateur des pratiques actuelles.

C. ELEMENTS THEORIQUES SUR LES ANALYSES STATISTIQUES

CLASSIQUES

Le paragraphe précédent a montré qu'une fois les données d'études de profil sensoriel obtenues, trois analyses statistiques sont le plus fréquemment utilisées :

- l'ACP, permettant d'obtenir une carte décrivant l'espace produit,
- l'ANOVA, dont l'objectif est de détecter des différences entre produits,
- les tests de comparaison multiple des moyennes des produits, permettant de détecter *a posteriori* des groupes de produits.

Nous allons nous intéresser principalement à l'ACP et à l'ANOVA. Le principe de chacune de ces analyses sera tout d'abord décrit, puis la théorie statistique sera présentée et illustrée par le jeu de données « cheeses » en guise d'exemple. Ce jeu de données est issu d'une étude de profil sensoriel sur des fromages dans laquelle 16 sujets ont évalué l'intensité de 12 descripteurs sur 6 fromages (A à F) en 3 répétitions. Ce jeu de données a été intégré à la plupart de nos bibliothèques R (ayant vocation à être utilisées par la communauté scientifique internationale) et les descripteurs sont donc libellés en anglais (Butter pour Beurre, Crumbly pour Croustillant, Dense pour Dense, Elastic pour Elastique, Fat pour Gras, Fruity pour Fruité, Hard pour Dur, Intense pour Intense, Milky pour Lacté, Salted pour Salé, Smooth pour Doux et Sticky pour Collant).

I. NOTATIONS

Les notations suivantes seront adoptées : les sujets seront indexés par $i = 1, \dots, I$, les produits par $j = 1, \dots, J$, les répétitions par $k = 1, \dots, K$ et les descripteurs par $d = 1, \dots, D$. La note obtenue pour un sujet i , un produit j , une répétition k et un descripteur d sera notée y_{ijk}^d .

Une matrice M avec N lignes et P colonnes sera notée $M(N, P)$ et pourra être définie par

$$M = (m_n^p)_{n=1..N, p=1..P} = \begin{pmatrix} m_1^1 & \cdots & m_1^P \\ \vdots & \ddots & \vdots \\ m_N^1 & \cdots & m_N^P \end{pmatrix} .$$

La transposée de M sera notée M' et correspond à : $M' = \begin{pmatrix} m_1^1 & \cdots & m_N^1 \\ \vdots & \ddots & \vdots \\ m_1^P & \cdots & m_N^P \end{pmatrix}$

II. ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

1. L'ACP DU TABLEAU DES MOYENNES DES PRODUITS

i. THEORIE DE L'ACP DU TABLEAU DES MOYENNES DES PRODUITS

Considérons $X(J, D)$ le tableau des moyennes des produits « colonne-centré » $X = (y_{j,d}^d)_{j=1..J; d=1..D}$ et $B(D, D)$ la matrice de covariance des produits.

$$B = \frac{1}{P-1} X'X$$

On peut choisir de standardiser ou non la matrice X , ce qui revient à donner le même poids à chaque descripteur. En analyse sensorielle, on utilise classiquement l'ACP non normée dite « de covariance ». En effet, de cette manière, les descripteurs étant le plus souvent notés sur la même échelle, un descripteur qui discrimine davantage les produits aura un poids plus important dans l'analyse. Inversement, un descripteur pour lequel il n'y a pas de différences entre les produits aura un poids moins important dans l'analyse..

L'analyse en composantes principales est basée sur la **diagonalisation** de B , qui permet de trouver les axes maximisant la variabilité entre les moyennes des produits, et donc de représenter le mieux possible les différences entre produits sur la carte d'ACP.

B est symétrique définie (ou semi-définie) positive et donc diagonalisable en base orthonormée. Il existe donc une base orthogonale de vecteurs propres $P = (e_1, \dots, e_D)$ et D valeurs propres positives $(\lambda_1 \dots \lambda_D)$ telle que $B = P \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \ddots & & 0 \\ 0 & 0 & & \lambda_D \end{pmatrix} P'$.

Pour exprimer un produit dans sa nouvelle base, il suffit alors de multiplier les coordonnées du produit dans l'ancienne base par la matrice de passage $P = (e_1, \dots, e_D)$. Par ailleurs, pour représenter les variables, on calcule la corrélation du vecteur des notes moyennes des produits dans l'ancienne base et le vecteur des notes moyennes des produit dans la nouvelle base. Dans la représentation graphique classique, deux graphiques sont construits :

- Le premier correspondant à la carte produit, représentant une section plane de l'espace produit (définie par deux axes) appelé plan principal,
- Le second correspondant aux corrélations entre les descripteurs et les deux axes choisis appelé cercle des corrélations.

Ces deux graphiques ne sont pas dans le même espace, et c'est pour cela qu'ils sont tracés sur des graphiques différents.

ii. EXEMPLE

Le tableau des moyennes (centrées par descripteur) du jeu de données « cheeses » est présenté Tableau 5.

	Butter	Crumbly	Dense	Elastic	Fat	Fruity	Hard	Intense	Milky	Salted	Smooth	Sticky
A	0.1587	1.1116	1.0259	1.0862	-0.6554	0.4186	1.3057	0.8616	0.1087	1.2381	-1.1248	-1.2345
B	-1.3483	2.1606	2.2967	1.8123	-1.4512	-1.2085	2.2557	-1.7998	-0.8694	-2.2036	-1.6342	-1.1491
C	0.8444	-1.0832	-1.0054	-0.9064	0.5915	0.6207	-1.1058	1.2606	0.5202	2.1464	0.7085	0.2498
D	-0.4462	0.5085	0.6707	0.5738	-0.2512	-0.4210	0.3880	-1.0946	-0.3955	-1.9245	-0.2592	-0.0887
E	0.8663	-1.5186	-1.5783	-1.8408	1.4207	0.9311	-1.7151	1.5335	0.4503	2.5536	1.3210	1.2081
F	0.2424	-1.1790	-1.4095	-0.7252	0.3457	-0.3408	-1.1286	-0.7613	0.1858	-1.8099	0.9887	1.01441

TABLEAU 5 : MOYENNES DES 6 PRODUITS SUR 11 DESCRIPTEURS

A partir de ce tableau (noté X), on obtient la carte d'ACP présentée Figure 3 :

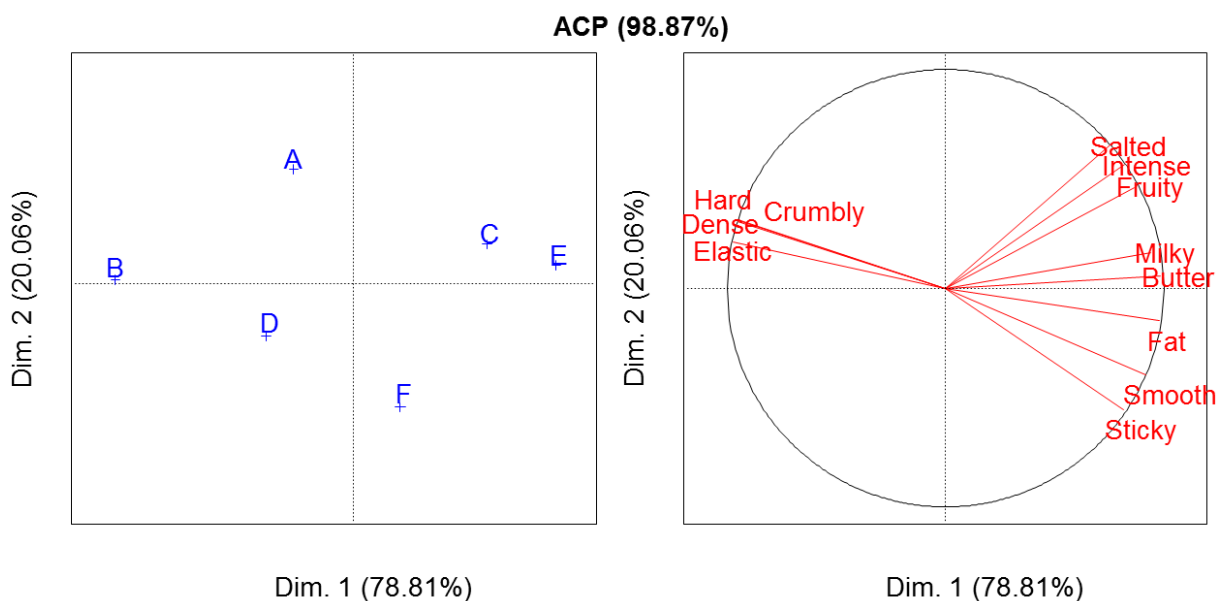


FIGURE 3 : L'ACP DES DONNEES DU TABLEAU 5

iii. INTERPRETATION

A gauche, on peut voir la carte produit. Les produits E et C d'une part et B d'autre part sont fortement discriminés sur le premier axe. A droite, sur la carte des descripteurs, on observe que ce premier axe est positivement corrélé aux descripteurs « Butter », « Milky », « Fat » ; et négativement corrélé à « Elastic », « Hard », « Crumbly » et « Dense ». On peut donc interpréter ce premier axe comme un axe opposant une texture dure à une sensation de gras et de beurre. Le second axe est plus difficilement interprétable. D'après la carte produit, E et C devraient avoir obtenu des notes plus élevées sur les descripteurs « Butter », « Milky » et « Fat » que B. On peut vérifier dans le tableau X que c'est bien le cas. F et A sont quant à eux discriminés par le deuxième axe.

Pour interpréter ce type de carte, il est nécessaire de **toujours revenir aux axes choisis**. Notamment, effectuer la **projection directe des produits sur un descripteur n'est pas correct** (même si c'est souvent ce qui est fait visuellement). Par exemple, si les produits se

projetent visuellement sur le descripteur « Hard » dans un certain ordre (ici B, A, D, F, C, E), on ne peut pas en déduire pour autant que les moyennes de ces produits pour ce descripteur « Hard » sont classées dans le même ordre (autrement dit que les produits sont, du plus dur (Hard) au moins dur, B, A, D, F, C, E). Pour cela il faudra utiliser le biplot (présenté dans le Chapitre I, C II.2.)

Le pourcentage à côté du nom de chaque axe indique le pourcentage de variabilité entre les produits expliqué par l'axe. Par conséquent, si la somme des pourcentages des deux axes choisis pour la carte est trop faible, cette carte seule ne représente pas bien les différences entre produits : elle doit être complétée par l'étude d'autres axes. Ce n'est pas le cas ici, puisque 98% de la variabilité est expliquée par le plan 1-2.

2. BIPLLOT

Jusqu'ici, l'ACP est constituée de deux graphiques, l'un correspondant aux produits, l'autre représentant le cercle de corrélation des descripteurs. Comme les deux graphiques ne représentent pas le même espace, il n'est pas possible de les superposer. Cependant, la tentation de représenter descripteurs et produits sur un même graphique est grande, et Gabriel (1971) propose une méthode pour le faire de manière correcte : le biplot.

i. PRINCIPE

Un biplot est une carte unique où sont représentés les produits (sous forme de points) et les descripteurs (sous forme de flèches). L'objectif du biplot est de pouvoir comparer directement les positions des descripteurs et de pouvoir projeter les points produits sur les flèches descripteurs.

ii. THEORIE

Il s'agit d'écrire la matrice X sous la forme d'un produit de deux matrices de points $A(J, 2)$ $B(D, 2)$ telle que $X = AB'$. Les points de A et B forment deux ensembles de points. L'un d'eux sera considéré comme les points du biplot (A), l'autre les vecteurs du biplot (B). Ainsi

$$x_{jd} = a_{j1}b_{d1} + a_{j2}b_{d2} = \langle a_j, b_d \rangle = \|a_j\| * \|b_d\| * \cos(a_j, b_d)$$

Considérons par exemple, le produit j et le descripteur d . Pour obtenir la note obtenue x_{jd} il suffit de mesurer la projection de j sur la flèche représentant le descripteur d puis de le multiplier par la norme du descripteur d (Figure 4).

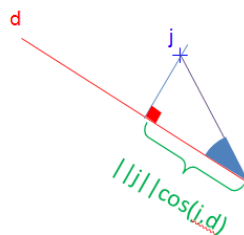


FIGURE 4 : ILLUSTRATION DE LA PROJECTION D'UN PRODUIT SUR UN ATTRIBUT DANS UN BIPLLOT

Pour exprimer A et B , on s'appuie sur **une décomposition en valeurs singulières (SVD)** (Golub & Kahan, 1965) qui permet d'écrire toute matrice rectangulaire X sous la forme :

$$X = U E V$$

où $E(J, D)$ une matrice rectangulaire diagonale (dont tous les coefficients e_{ij} sont nuls pour $i \neq j$) de même rang que X et $U(J, J)$ et $V(D, D)$ deux matrices orthonormées.

On décompose ensuite $X = U E^{1-\alpha} E^\alpha V$. Puis on pose $A = U E^\alpha$ et $B = E^{1-\alpha} V$. On remarque que A et B n'ont pas que deux colonnes (correspondant aux abscisses et aux ordonnées des points). Cependant, en ne conservant que les deux premières colonnes de la matrice A et B , on espère arriver à une bonne approximation de X .

Si $\alpha = 1$, alors $A = U E$ et $B = V$. C'est ce qu'on appelle **le biplot des distances** :

La matrice de variance estimée est $B = X'X / (n - 1)$.

$$B = V' E' U' U E V = V' E E V \text{ (car } U \text{ est orthonormale).}$$

On remarque alors que V correspond aux vecteurs propres de B et que $E'E$ correspond à la matrice diagonale des valeurs propres (ordonnées par ordre décroissant). Les coordonnées des produits en ACP classique étaient obtenus par $XV' = U E V V' = U E = A$.

Les coordonnées des produits de l'ACP classique sont donc strictement égales à celles obtenues pour le biplot : le biplot conserve la « configuration des points de l'ACP », en revanche les coordonnées des descripteurs changent.

Si $\alpha = 0$, le biplot est appelé « biplot des corrélations » : il conserve les corrélations et modifie les positions des produits.

iii. EXEMPLE SUR LE JEU DE DONNEES CHEESES

Le biplot ($\alpha = 1$) effectué sur les données de l'exemple donne le résultat présenté Figure 5.

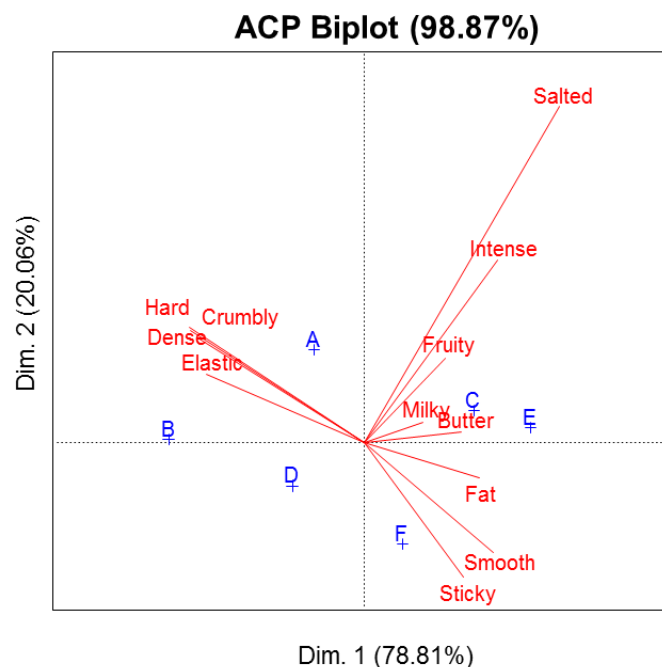


FIGURE 5 : BIPLLOT DES DONNEES DE L'EXEMPLE ($\alpha = 1$)

iv. INTERPRETATION

Sur la Figure 5, les positions des produits sont exactement similaires à celle de l'ACP classique (Figure 3). En revanche, la projection des produits sur les descripteurs est à présent autorisée, et ainsi on peut supposer qu'on a, du plus « Hard » au moins « Hard », B, A, D, C, F, E (confirmé par le tableau de données). Les longueurs des flèches représentant les descripteurs varient et sont interprétées comme « taille des différences entre les produits ». Par exemple, on peut voir que la flèche de « Salted » est beaucoup plus longue que les autres. Cela signifie que les différences entre les produits ont été plus grandes pour « Salted », ce qu'on peut retrouver dans le tableau des données (Tableau 7).

L'ACP est le plus souvent représentée sans l'option biplot. Cependant, malgré la séparation des deux graphiques, elle est souvent mal interprétée, car l'utilisateur a tendance à projeter inconsciemment les positions des produits sur les descripteurs sans passer par l'interprétation des axes factoriels. Par conséquent, le biplot, même s'il est basé sur une théorie plus complexe (décomposition en valeurs singulières) offre une carte plus facilement interprétable par l'utilisateur.

III. ANOVA ET DETECTION DE DIFFERENCES UNIDIMENSIONNELLES ENTRE LES PRODUITS

1. PRINCIPE

L'Analyse de la variance (ANOVA) est un **traitement statistique** qui s'applique lorsque l'on veut savoir s'il existe **un effet d'un ou plusieurs facteurs qualitatifs sur une variable quantitative**. Dans le cas d'un profil sensoriel, elle est le plus souvent utilisée pour savoir s'il existe des différences entre les produits.

L'ANOVA est basée sur le **choix d'un modèle** qui permet d'expliquer la note par l'addition des effets de facteurs qualitatifs (produit, sujet, répétition...). A partir de ce modèle, on décompose la variabilité totale des observations **en somme de variabilités dues aux effets des différents facteurs**. Si la variance spécifique à un effet « produit » est statistiquement plus élevée qu'une **variance « référence »**, on pourra déduire qu'il y a un effet significatif et conclure à des différences entre les produits.

2. THEORIE

i. CHOIX DU MODELE

Le modèle classiquement utilisé en analyse sensorielle est **le modèle à deux facteurs** (produit et sujet) suivant :

$$y_{ijk}^d = m^d + s_i^d + p_j^d + d_{ij}^d + \varepsilon_{ijk}^d$$

Où y_{ijk}^d représente la note du sujet i au produit j pour le descripteur d à la répétition k ; m^d : moyenne générale, p_j^d : effet produit, s_i^d : effet sujet, d_{ij}^d l'interaction produit par sujet, et ε_{ijk}^d le terme d'erreur résiduelle, avec $\varepsilon_{ijk}^d \sim N(0, \sigma^2)$ indépendantes.

Selon Næs and Langsrud (1998), en analyse sensorielle, considérer l'effet sujet comme fixe revient à affirmer que ces sujets sont délibérément choisis par l'animateur de panel. Au contraire, considérer un effet sujet comme aléatoire revient à voir les sujets comme un échantillon représentatif d'une population. Par conséquent, le modèle approprié aux données de profil sensoriel serait un modèle avec effet produit fixe et effet sujet aléatoire. Ce modèle est alors appelé **modèle mixte** (comprenant différents types d'effets) et s'écrit de la même manière que le modèle à deux facteurs, mais avec $d_{ij}^d \sim N(0, \sigma_d^2)$ et $s_i^d \sim N(0, \sigma_j^2)$ (effet sujet et interaction aléatoires).

Le choix du type d'effet impacte directement les calculs de l'ANOVA.

ii. DECOMPOSITION DES SOMMES DES CARRÉS

m^d étant estimée par $y_{...}$; p_j^d par $y_{.j} - y_{...}$, s_i^d par $y_{i.} - y_{...}$, d_{ij}^d par $y_{ij.} - y_{.j} - y_{i.} + y_{...}$, on note les variabilités dues à l'effet produit, sujet, produit*sujet (interaction), de l'erreur et la variabilité totale par les sommes des carrés suivantes :

$$SS_{produit} = I * K * \sum_j (y_{.j} - y_{...})^2$$

$$SS_{sujet} = J * K * \sum_{i=1}^I (y_{i.} - y_{...})^2$$

$$SS_{produit*sujet} = K * \sum_{i=1}^I \sum_{j=1}^J (y_{ij.} - y_{.j} - y_{i.} + y_{...})^2$$

$$SS_{erreur} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{ij.})^2$$

$$SS_{total} = \sum_{i,j,k} (y_{ijk} - y_{...})^2$$

Lorsque les panélistes ont goûté tous les produits le même nombre de fois (et les ont par conséquent notés sur l'ensemble des descripteurs), on dit que le **jeu de données est équilibré**. Alors, on peut montrer que

$$SS_{total} = SS_{produit} + SS_{sujet} + SS_{produit*sujet} + SS_{erreur}$$

C'est la **décomposition des sommes des carrés**. Cette égalité n'est plus valable lorsque les jeux de données ne sont pas équilibrés, et l'ANOVA nécessite alors des ajustements (type I, II ou III de SAS). Dans notre travail, nous nous restreignons généralement aux jeux de données équilibrés pour effectuer des ANOVA.

iii. CHOIX DE LA VARIANCE REFERENCE

Pour rechercher si l'effet produit est significatif, on pose $MS_{produit} = \frac{SS_{produit}}{J-1}$ qui correspond au carré de l'écart moyen entre les produits (ou variance des points moyens), et on cherche à la comparer à une variance « référence » à l'aide du **test de Fisher**.

Dans le cas où l'on considère que l'effet sujet est fixe, on va comparer $MS_{produit}$ à MS_{erreur} : le niveau de non-répétabilité des sujets. En effet, le terme générique de l'erreur s'écrit $(y_{ijk} - y_{ij.})$ et correspond donc, pour un produit et un sujet fixés, à l'écart à la moyenne des notes attribuées aux différentes répétitions.

Dans le cas où le modèle choisi est mixte (l'effet sujet est aléatoire), on compare $MS_{produit}$ à $MS_{produit*sujet}$, le désaccord moyen entre les juges. En effet, le terme générique de l'interaction s'écrit :

$$(y_{ij.} - y_{.j.} - y_{i..} + y_{...}) = (y_{ij.} - y_{i..}) - (y_{.j.} - y_{...})$$

et correspond donc à la différence entre les notes centrées du panéliste et les notes centrées du panel.

iv. TEST STATISTIQUE

On se place sous l'hypothèse H_0 : « les moyennes théoriques des produits sont égales ».

$F = MS_{produit}/MS_{produit*sujet}$ est appelée statistique de Fisher. Sous H_0 , F suit une loi de Fisher de degrés de liberté $(J - 1)$ et $(I - 1)(J - 1)$, sous réserve de normalité des résidus, d'homoscédasticité (homogénéité) des variances et d'indépendance des échantillons. Cette statistique sera comparée avec la distribution de Fisher théorique et une p-value sera obtenue. Si cette p-value est inférieure à un seuil alpha (risque de première espèce) préalablement défini (en général 0.05), on rejette H_0 , et l'effet du facteur est significatif : certaines moyennes ne sont pas égales entre elles.

v. ECRITURE DU TABLEAU DE L'ANOVA

Dans le cas d'un modèle à deux facteurs (produit et sujet) avec effet sujet aléatoire, on obtient alors le tableau d'ANOVA suivant (Tableau 6).

Variable	Degré de liberté	de SS	MS	F	Pval
Produit	$J - 1$	$SS_{produit}$	$\frac{SS_{produit}}{J - 1}$	$\frac{MS_{produit}}{MS_{produit*sujet}}$	
Sujet	$I - 1$	SS_{sujet}	$\frac{SS_{sujet}}{I - 1}$	$\frac{MS_{sujet}}{MS_{produit*sujet}}$	
Interaction Produit*Sujet	$(I - 1)(J - 1)$	$SS_{produit*sujet}$	$\frac{SS_{produit*sujet}}{(J - 1)(I - 1)}$	$\frac{MS_{produit*sujet}}{MS_{erreur}}$	
Résidus (ou Erreur)	$IJ(K - 1)$	SS_{erreur}	$\frac{SS_{erreur}}{IJ(K - 1)}$		

TABLEAU 6 : TABLEAU D'ANOVA D'UN MODELE A DEUX FACTEURS A EFFET SUJET ALEATOIRE

La première ligne permet de déterminer s'il y a un effet produit, la seconde un effet sujet et la troisième un effet de l'interaction.

3. EXEMPLE

Pour le jeu de données « Cheeses » et le descripteur « Salted », on obtient les résultats suivants (Tableau 7).

Variable	Degré de liberté	SS	MS	F	P-value
Produit	5	1175.81	235.162	53.555	< 0.001
Sujet	15	572.93	38.195	8.698	< 0.001
Interaction Produit*Sujet	75	329.32	4.391	2.7714	< 0.001
Résidus (ou Erreur)	192	304.21	1.584		

TABLEAU 7 : TABLEAU D'ANOVA DU JEU DE DONNEES CHEESES

4. INTERPRETATION

Cela signifie qu'en prenant un risque $\alpha = 0.05$, l'effet produit (< 0.05) est significatif dans ce jeu de données. Autrement dit, il existe au moins un produit dont la moyenne est différente de celle d'un autre pour le descripteur Salted. Les effets sujet et de l'interaction sont également significatifs.

IV. PROBLEMES THEORIQUES DE L'ACP ET DE L'ANOVA

L'ACP et l'ANOVA sont les méthodes statistiques les plus utilisées en analyse de profil sensoriel mais ont néanmoins certaines limites.

Tout d'abord, l'ACP est **basée sur le tableau des moyennes des produits (B)**. Cependant, une moyenne de 5 obtenue pour un produit peut être obtenue parce que tous les sujets ont noté 5, ou parce que la moitié des sujets a noté 0 et l'autre moitié 10. Dans le premier cas, aucun désaccord n'est observé, dans l'autre le désaccord est maximal. L'ACP ne fait aucune différence entre ces deux situations, et représentera la même carte, *que les sujets soient d'accord sur les moyennes attribuées ou non*. Plus précisément, **l'ACP ne prend pas en compte l'interaction entre les sujets dans sa représentation graphique**. Cela présente un problème dans la mesure où la carte tracée n'a aucun modèle statistique de notes sous-jacent. Ce point sera développé dans le Chapitre 3 I. 1. De plus, en profil sensoriel, il est courant d'avoir plus de descripteurs que de produits. Dans ce cas, on a plus de variables que d'observations pour estimer la matrice de covariance, ce qui n'est pas recommandé. Ces limites de l'ACP constituent un réel problème théorique pour la production de cartes en profil sensoriel.

Ensuite, le **modèle** utilisé dans l'ANOVA à deux facteurs prend bien en compte l'effet sujet (certains sujets notent plus ou moins haut sur l'échelle), mais **ne prend pas en compte le fait que ces sujets dispersent plus ou moins leur notes sur l'échelle –effet de scaling–**. L'effet de scaling est jusqu'ici considéré comme une source de désaccord entre les sujets, alors qu'il ne s'agit que d'un effet d'utilisation de l'échelle. Pour corriger cet effet, Brockhoff, Schlich, and Skovgaard (2015) proposent un nouveau modèle : le Mixed Assessor Model. Ce modèle, présenté dans le Chapitre 3. A. 2, permet de prendre cet effet en compte et de le corriger.

D. OBJECTIFS ET PLAN DE CETTE THESE

Dans cette introduction, l'analyse sensorielle a été décrite comme une discipline utilisant des sujets humains en tant qu'outil de mesure. Le fonctionnement des 5 sens a été décrit puis nous avons mis en évidence diverses sources de variabilité (physiologiques ou psychologiques) pouvant avoir un impact sur la mesure sensorielle. Un état de l'art du profil sensoriel, basé sur une étude bibliographique a également conduit à une documentation des pratiques actuelles du profil sensoriel. Enfin, les analyses statistiques les plus utilisées pour l'analyse du profil sensoriel - Analyse en Composantes Principales (ACP) et Analyse de la Variance (ANOVA) - ont été décrites et leurs limites mises en évidence :

- l'ACP ne prend pas en compte le fait que les panélistes soient plus ou moins d'accord sur les notes attribuées aux produits,
- le modèle classique de l'ANOVA ne prend pas en compte le fait que certains panélistes dispersent leurs notes plus que d'autres.

L'objectif de ce travail de thèse est de **revisiter l'analyse de profil sensoriel par une approche base de données**. Les chapitres suivants (2 à 5) vont nous permettre de réaliser cet objectif.

Dans le chapitre 2, nous présenterons la Sensobase (www.sensobase.fr), une base de jeux de données de profil sensoriel et nous la documenterons (nombre de sujets, de produits, de descripteurs par jeu de données...). Puis, une méthodologie pour traiter simultanément un grand nombre de jeux de données sera mise en place. Elle sera inspirée des méta-analyses déjà effectuées dans le domaine de l'analyse sensorielle, mais aussi de celles réalisées en essais cliniques. La méthodologie de méta-analyses sera ensuite directement appliquée sur la SensoBase pour répondre à différentes questions (pourcentage de zéros dans les notes attribuées en profil sensoriel, effet de l'utilisation d'un modèle mixte plutôt qu'un modèle à effet fixe dans l'ANOVA...).

Le troisième chapitre proposera des solutions pour résoudre les problèmes de l'ACP et de l'ANOVA, notamment en ce qui concerne les performances du panel et la cartographie des produits. Nous insisterons particulièrement sur la notion de scaling (tendance des panélistes à disperser plus ou moins leurs notes que le panel sur l'échelle de notation d'un descripteur). De nouvelles méthodes de calcul de performances unidimensionnelles prenant le scaling en compte seront mises en place. Un outil intégrant ces performances à destination des animateurs de panel sera également proposé : la table MAM-CAP (Peltier, Brockhoff, Visalli, & Schlich, 2014). Le scaling sera ensuite étudié de manière multidimensionnelle (tous descripteurs confondus) et un modèle multidimensionnel associé sera proposé : le MultiMAM (Peltier, Visalli, & Schlich, 2015e). Il en résultera une décomposition de la dispersion des notes en deux composantes : la première indépendante du descripteur noté (tendance psychologique à sur-disperser ses notes), et la seconde décrivant la sensibilité du sujet à un descripteur fixé (Peltier, Visalli, & Schlich, 2015d).

Par ailleurs, la Canonical Variate Analysis (CVA), une carte prenant en compte la variabilité des sujets autour des moyennes des produits sera proposée comme une alternative à l'ACP. Cette carte sera mise à disposition des utilisateurs à l'aide d'une librairie R dédiée (Peltier et al., 2015e). La CVA sera ensuite adaptée au MultiMAM (MAM-CVA), ce qui permettra de tracer une carte prenant en compte l'effet de scaling.

Le quatrième chapitre appliquera les méthodes présentées dans le chapitre 3 sur un grand nombre de jeux de données. Ainsi, nous étudierons les liens entre performances (uni et multidimensionnelles) et caractéristiques des panélistes (âge, genre...). Puis la CVA sera comparée à l'ACP (Peltier, Visalli, & Schlich, 2015b). Pour cela, une méthodologie de comparaison de cartes sera mise en place. Cette méthodologie sera réutilisée pour comparer d'autres cartes (MAM-CVA et ACP par exemple). Enfin, la question de l'influence du nombre de répétitions sur les résultats (Peltier, Visalli, & Schlich, 2015c) sera abordée à l'aide des méthodes mises en place.

Finalement, lors de la discussion, nous synthétiserons les résultats obtenus et les mettrons en regard avec l'étude bibliographique du profil sensoriel. Cela nous permettra d'aboutir à des recommandations sur l'analyse des données, mais aussi sur la mise en place du

protocole de profil sensoriel. La Figure 31 synthétise l'ensemble du travail effectué et le lecteur pourra s'y référer.

CHAPITRE 2 :

MATERIEL ET METHODES

A. SENSOBASE

I. PRESENTATION

1. DEFINITION ET HISTORIQUE

La SensoBase est une base de données regroupant des études de profil sensoriel. Créée en 2006 au Centre des Sciences du Goût et de l'Alimentation par Nicolas Pineau et Pascal Schlich (Pineau, 2006), elle a été conçue afin d'accumuler des données de profil sensoriel provenant de divers partenaires (publics ou privés), avec pour objectifs :

- de **documenter les pratiques de l'évaluation sensorielle** ; par exemple étudier les distributions des nombres de produits, sujets, descripteurs dans une étude de profil sensoriel,
- de **suggérer des améliorations des pratiques de l'évaluation sensorielle** ; par exemple, étudier l'impact des caractéristiques individuelles des panélistes, de la modalité sensorielle du descripteur ou du type de produit sur la perception,
- de **comparer des méthodes d'analyse des données sensorielles** à partir d'un grand nombre de jeux de données.

2. STRUCTURE DE 2006 A 2015

A partir de 2006, la SensoBase était accessible via Internet (www.sensobase.fr). Les fournisseurs s'inscrivaient en ligne, puis déposaient leurs jeux de données de profil afin d'obtenir en retour des analyses statistiques (performances du panel, cartes...). Pour cela, ils présentaient un fichier Excel avec leurs données et divers renseignements sur l'étude (caractéristiques des sujets, conditions d'expérience...). Les données étaient stockées dans une base MySQL.

3. EVOLUTION DE LA SENSOBASE

Début 2015, le site www.sensobase.fr a été fermé et la base de données a été transférée dans « SensoBaseV2 » accessible via le logiciel TimeSens, avec une interface beaucoup plus ergonomique (les données acquises dans TimeSens peuvent être importées dans SensoBase en 3 clics). Cette nouvelle base permet de collecter également les jeux de données hédoniques (tests de préférence) et temporels (DTS). Elle a pour vocation de **permettre à l'utilisateur de TimeSens non seulement de collecter ses propres jeux de données, mais aussi de les rassembler dans sa propre base et de pouvoir ainsi analyser « en bloc »**. Certaines des méta-analyses développées dans cette thèse seront proposées ultérieurement à l'utilisateur de TimeSens afin qu'il puisse les appliquer à ses propres données. Les parrains de cette thèse ont abondé la SensoBase. Ainsi, ils ont fourni plus de 150 jeux de données que nous avons intégrés dans la SensoBase.

4. SOUS ENSEMBLE DE DONNEES UTILISE

Puisque la SensoBase est alimentée de manière continue, une **version stable** de la SensoBase contenant 1168 jeux de données **a été enregistrée en Décembre 2013** afin de conserver la même base de données pour l'ensemble de nos travaux. Tous les jeux de données entrés dans la SensoBase après cette date ne seront pas pris en compte. Dorénavant, le terme SensoBase fera référence à la base de données enregistrée avec 1168 jeux de données.

II. DESCRIPTION DE LA SENSOBASE

1. CARACTERISTIQUES DES JEUX DE DONNEES

Les 1168 jeux de données de la SensoBase référencent 5 696 sujets, 7 069 produits, 19 447 descripteurs pour 5 411 462 observations enregistrées. Les distributions des nombres de sujets, produits et descripteurs sont données dans la Figure 6 :

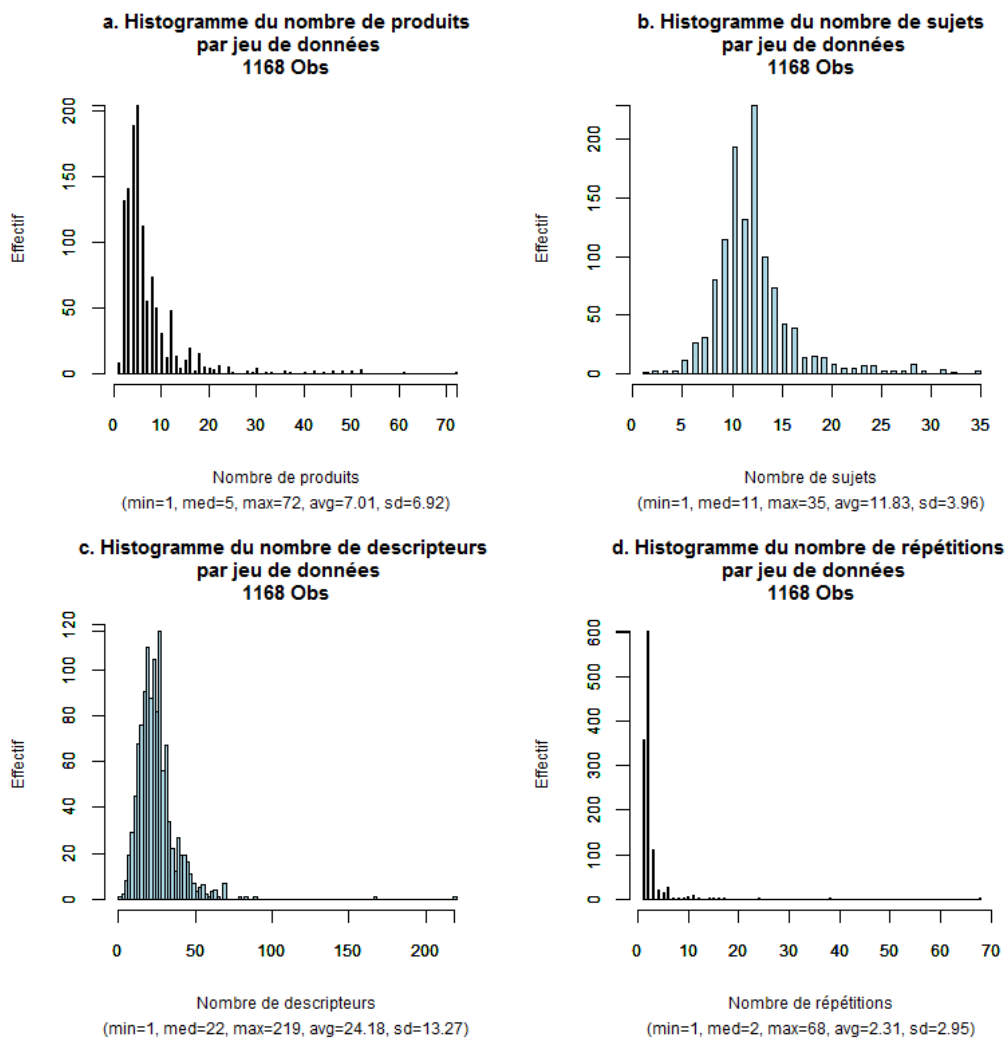


FIGURE 6 : DISTRIBUTION DES CARACTERISTIQUES DE LA SENSOBASE

Ces distributions montrent que le jeu de données « médian » comprend 5 produits, 11 sujets et 22 descripteurs. De plus, on peut observer que certains jeux de données ont un grand nombre de descripteurs (>100) ou de produits (>20). La distribution du nombre de répétitions montre également un jeu de données atypique avec 68 répétitions. Ces jeux atypiques ne seront pas pris en compte lors des analyses.

Le Tableau 8 montre que plus de deux tiers des jeux de données ont deux répétitions ou plus (« 1 » répétition signifie qu'il n'y a pas eu de répétitions).

Nombre de répétitions	1	2	3	4	5	6	7+	Total
Nombre de jeux de données	360	603	111	20	13	25	36	1168

TABLEAU 8 : NOMBRE DE REPETITIONS DANS LA SENSOBASE

La distribution de la variable quantitative « score » (traduisant la note donnée à un descripteur pour un sujet/produit/répétition ramenée entre 0 et 10) est présentée Figure 7.

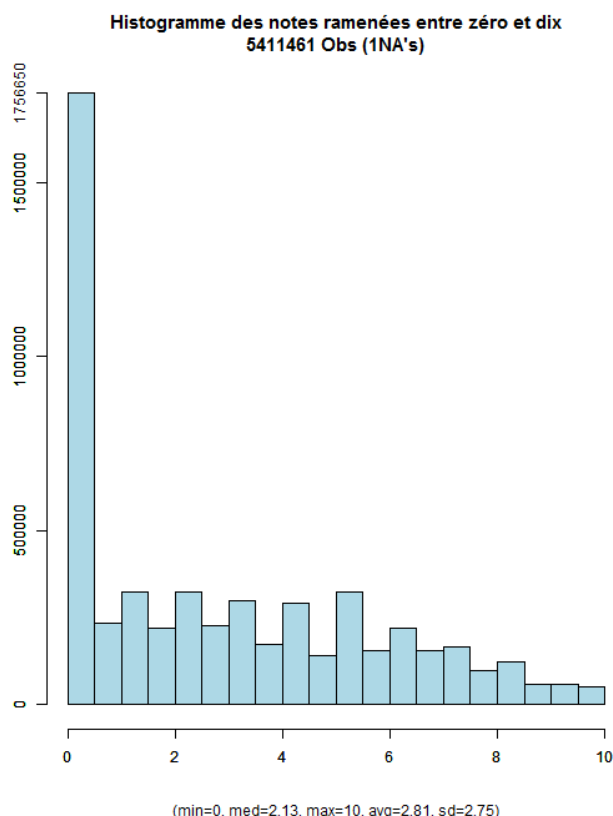


FIGURE 7 : HISTOGRAMME DES NOTES RAMENEES ENTRE 0 ET 10

26.86 % des observations sont égales à zéro. Cette observation fera l'objet d'une méta-analyse à la fin de ce chapitre (D.II, p 53). On peut supposer que ces zéros sont dus principalement à des descripteurs d'odeur, qui ne sont pas perçus par l'ensemble des sujets. La forte proportion de zéros peut également s'expliquer par le fait que certains descripteurs intégrés dans l'étude pourraient ne pas être pertinents pour tous les produits de cette étude.

2. CARACTERISTIQUES DES SUJETS

En plus des données brutes, la SensoBase contient également des caractéristiques décrivant les sujets (genre, âge, niveau d'étude, consommation de tabac), les produits (catégorie du produit), les descripteurs (type d'échelle, modalité sensorielle), et les fournisseurs (pays d'origine). Notons que certains sujets/produits/descripteurs impliqués dans plusieurs études sous des identifiants différents peuvent avoir été comptés plusieurs fois.

Certains sujets ont participé à plusieurs études (avec le même identifiant). En réalité, 13 821 couples sujet*jeu de données ont été observés (sur les 5 696 sujets étudiés dans la SensoBase). Ainsi, au total, 18% des sujets ont participé à plus d'une étude (Tableau 9).

Nombre de participations	1	2 à 5	6 à 10	11 et 49	50 à 100	+ de 100	Total
Nombre de sujets	4 687	679	139	164	22	5	5 696

TABLEAU 9 : NOMBRE DE PARTICIPATIONS PAR SUJET

i. GENRE

L'ensemble des panélistes de la SensoBase est constitué d'environ 70% de femmes et de 30% d'hommes (Tableau 10).

Genre	Effectif	Pourcentage
Femme	3 974	69.77 %
Hommes	1 517	26.63 %
Non renseigné	205	3.60 %
Total	5 696	100 %

TABLEAU 10 : REPARTITION DES SUJETS EN FONCTION DU GENRE DANS LA SENSOBASE

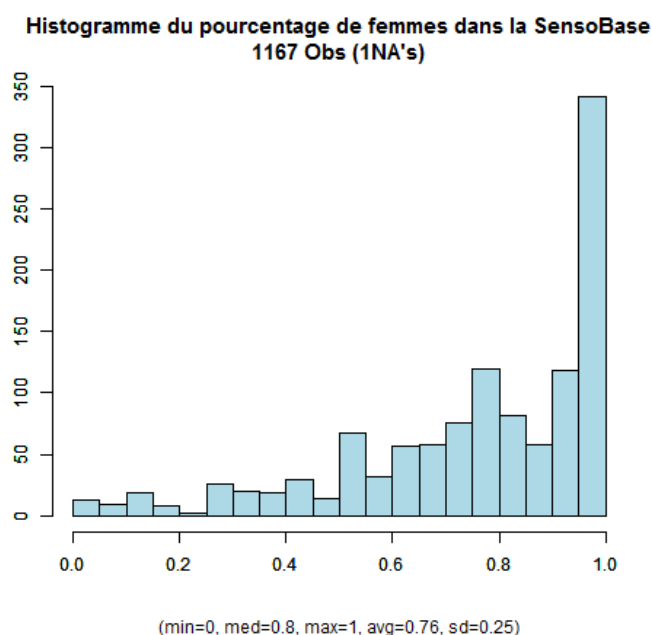


FIGURE 8 : HISTOGRAMME DU POURCENTAGE DE FEMMES DANS LA SENSOBASE

Ce déséquilibre est courant dans les études de profil sensoriel et peut s'expliquer soit par le fait que les femmes sont plus disponibles que les hommes pour participer aux séances, soit qu'elles souhaitent davantage participer aux études sensorielles. Selon les chiffres de l'INSEE au premier janvier 2014, la proportion de femmes dans la population française est de 51.55 %. Par conséquent, les échantillons ne sont pas souvent représentatifs de la population française. La distribution du pourcentage de femmes par jeu de données est donnée par la Figure 8. Elle montre que moins de 200 jeux de données sur 1 167 ont entre 40 et 60% de femmes dans leur panel. Par ailleurs, 28.96 % des jeux de données ont 100 % de femmes.

ii. AGE

L'âge au moment de l'étude est calculé à l'aide d'une requête SQL basée sur la date de naissance du sujet et la date de l'étude. On obtient 13 821 observations, dont 3 sont des âges négatifs et 22 des âges supérieurs à 100 ans. Ces valeurs aberrantes seront remplacées par des valeurs manquantes (NA).

L'âge des sujets varie de 0 à 93 ans et l'âge moyen et médian est 48 ans (Figure 9).

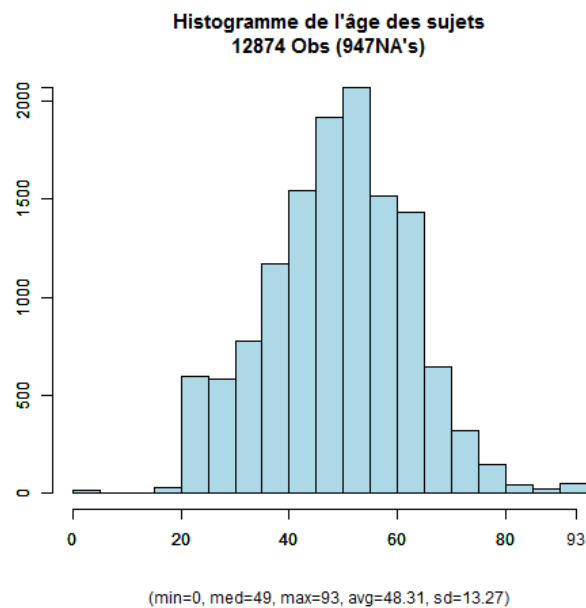


FIGURE 9 : DISTRIBUTION DE L'AGE DES SUJETS DANS LA SENSOBASE

Cette variable Age peut être divisée en classes (Tableau 11).

	Effectif	Pourcentage
[0 ; 20 [113	0.82 %
[20 ; 30 [1 340	9.70 %
[30 ; 40 [1 992	14.41 %
[40 ; 50 [3 568	25.82 %
[50 ; 60 [3 489	25.24 %
[60 ; 70 [1 866	13.50 %
[70 ; 100 [521	3.66 %
Non renseigné	932	6.85 %
Total	13 821	100 %

TABLEAU 11 : REPARTITION DE L'AGE DES SUJETS DANS LA SENSOBASE

iii. NIVEAU D'ETUDE

Le Tableau 12 montre que les participants aux analyses sensorielles ont généralement un niveau d'études supérieures. Ce résultat est à moduler puisqu'on a 58% de données manquantes.

Niveau Etudes	Effectif	Pourcentage
Donnée inconnue ou NA	953	58.22 %
Etudes primaires (avant collège)	27	1.65 %
Etudes secondaires (collège et lycée)	151	9.22 %
Etudes supérieures	506	30.91 %
Total	1637	100 %

TABLEAU 12 : REPARTITION DU NIVEAU D'ETUDE DES SUJETS DANS LA SENSOBASE

iv. FUMEUR

Parmi les sujets pour lesquels la variable est renseignée, on dénombre 103 fumeurs contre 740 non-fumeurs et 107 anciens fumeurs (Tableau 13). En 2000, le baromètre santé de l'INPS dénombrait 30 % de français fumeurs et 4 % de fumeurs occasionnels en France.

Fumeur	Effectif	Pourcentage
Non fumeur	740	45.20 %
Fumeur	103	6.29 %
Ancien fumeur	107	6.54 %
NA	687	41.97 %
Total	1 637	100 %

TABLEAU 13 : REPARTITION DES FUMEURS DANS LA SENSOBASE

La distribution du pourcentage de fumeurs dans les jeux de données montre que 69.75 % des jeux de données ne contiennent pas de fumeurs (Figure 10). Là encore, le résultat est à moduler en raison du fort taux de valeurs manquantes (42%).

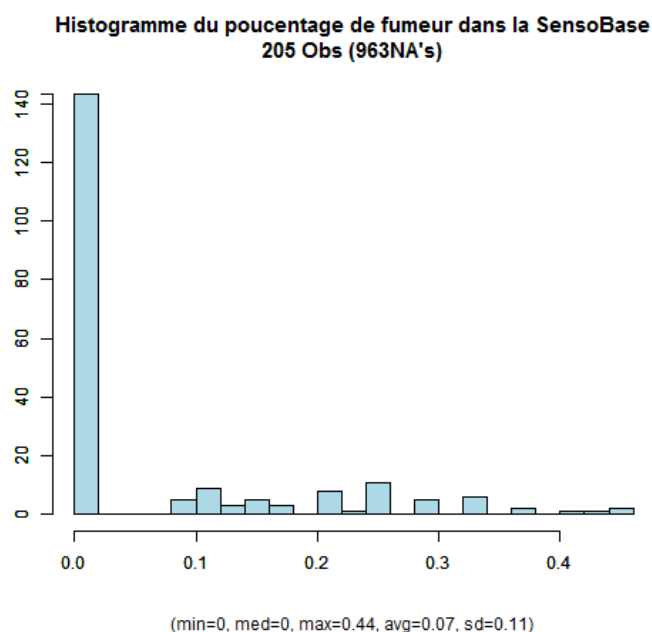


FIGURE 10 : HISTOGRAMME DU POURCENTAGE DE FUMEURS DANS LA SENSOBASE

3. CARACTERISTIQUES DES PRODUITS DE LA SENSOBASE

Les catégories de produits les plus étudiées sont les yaourts, petit suisses, fromage blanc et fromages (Tableau 14).

Type de produit	Effectif	Pourcentage
Produits laitiers (yaourt, lait, fromages...)	3091	43.73 %
Boissons alcoolisées	840	11.88 %
Sauces et condiments	781	11.05 %
Plats préparés	349	4.94 %
Sucreries	333	4.71 %
Charcuteries	323	4.57 %
Boissons non alcoolisées	226	3.20 %
Fruits	211	2.99 %
Produits céréaliers (pain, biscuits...)	203	2.87 %
Huiles et graisses	149	2.11 %
Autres produits	147	2.07 %
Légumes	126	1.78 %
Poissons	115	1.63 %
Féculents (pomme de terre et légumes secs...)	108	1.53 %
Viandes	63	0.89 %
Œufs et dérivés	3	0.04 %
Total	7 068	100 %

TABLEAU 14 : NOMBRE DE PRODUITS PAR CATEGORIE DE PRODUIT

4. CARACTERISTIQUES DES DESCRIPTEURS DE LA SENSOBASE

i. ECHELLE UTILISEE

Les deux principaux types d'échelles sont les échelles discrètes et les échelles continues. La forme des échelles peut varier. Ainsi, les échelles continues peuvent être sous forme rectangulaires (placement d'une barre dans un rectangle) ou linéaires (placement d'une note sur une ligne). Une échelle linéaire peut être fermée (bornes de l'échelle délimitées) ou ouverte (échelle non bornée). Les principales échelles utilisées dans la SensoBase sont les échelles continues linéaires fermées et les échelles discrètes (Tableau 15).

Echelle utilisée	Effectif	Pourcentage
Echelle inconnue ou non renseignée	5 271	27.11 %
Echelle continue rectangulaire	493	2.54 %
Echelles continues linéaires fermées	7 670	39.44 %
Echelles continues linéaires ouvertes	502	2.58 %
Echelles discrètes	5 458	28.07 %
Autre	51	0.27 %
Total	19 445	100 %

TABLEAU 15 : ECHELLES UTILISEES DANS LA SENSOBASE

ii. MODALITE SENSORIELLE

Les modalités sensorielles les plus présentes dans la SensoBase sont les arômes et les odeurs (Tableau 16).

Famille de descripteur	Effectif	Pourcentage
Non renseignée	92	0.47 %
Apparence	2 518	12.95 %
Odeur	4 045	20.80 %
Saveur	3 129	16.09 %
Texture	3 509	18.05 %
Arôme/flaveur	4 531	23.30 %
Arrière-gout	1 006	5.17 %
Trigéminal	611	3.14 %
Hédonique	4	0.02 %
Total	19 445	100 %

TABLEAU 16 : FAMILLES DE DESCRIPTEURS DANS LA SENSOBASE

On peut noter qu'il ne devrait théoriquement pas y avoir de descripteurs hédoniques dans la SensoBase puisqu'il s'agit de profils sensoriels : cette étude est atypique.

5. CARACTERISTIQUES DES FOURNISSEURS

La SensoBase a été abondée par 66 fournisseurs. 38 d'entre eux sont français et 16 étrangers. On distingue les petits fournisseurs (moins de 10 jeux de données), les fournisseurs moyens (entre 10 et 50 jeux de données) et les gros fournisseurs (plus de 50 jeux de données (Tableau 17)).

On observe également que certaines entreprises sont « sur-représentées ». Nous aurions pu, dans le reste de nos travaux de thèse, limiter le nombre de jeux de données par entreprise (comme cela avait été le cas dans Pineau and Schlich (2015)), mais nous avons préféré conserver le nombre maximal de jeux de données.

	Effectif	Pourcentage
Petits fournisseurs (<10 jeux de données)	47	71.21 %
Fournisseurs moyens (entre 10 et 50 jeux de données)	13	19.70 %
Gros fournisseurs (plus de 50 jeux de données)	6	9.09 %
Total	66	100 %

TABLEAU 17 : DISTRIBUTION DES FOURNISSEURS DE LA SENSOBASE SELON LE NOMBRE DE JEUX DE DONNEES FOURNIS

6. JEUX DE DONNEES NON EQUILIBRES

Un jeu de données est **complet** quand tous les sujets ont noté sur l'ensemble des descripteurs tous les produits au moins une fois. De plus, il est **équilibré** quand tous les sujets ont noté l'ensemble des descripteurs pour tous les produits le même nombre de fois. Classiquement, on détecte si un jeu de données est équilibré en vérifiant que le nombre d'observations est égal au nombre de produits x nombre de répétitions x nombre de sujets x nombre de descripteurs. Par ailleurs, l'équilibrage ou la complétude des données peuvent également être étudiés par descripteur lorsqu'on s'intéresse à l'analyse unidimensionnelle des données.

Sur les 1168 jeux de données de la SensoBase, on a observé que 113 jeux de données avaient des données dupliquées (plusieurs notes pour une observation sujet x produit x descripteur x répétition x session fixée. Sur les **1 055** (90.33 %) jeux de données sans données dupliquées, 952 (90.24 %) sont complets et 742 (70.33 %) sont équilibrés. Cependant, parmi ces jeux de données équilibrés, certains utilisateurs ont confondu session (date de la séance) et répétition (nombre de fois où le produit a été goûté) : selon eux, un produit était évalué par un sujet lors de plusieurs sessions, mais au cours de la même répétition, ce qui est impossible ! Cette confusion a eu lieu pour 207 jeux de données. Au total, **seuls 531 (45.46 %) des jeux de données sont vraiment équilibrés.**

7. SYNTHÈSE

La SensoBase est une base de données **unique au monde** qui regroupe 1 168 jeux de données de profil sensoriel et qui continue d'être alimentée dans TimeSens. Les problèmes techniques suivants ont été rencontrés :

- Certaines variables ne sont **pas remplies** ou **mal renseignées** : lignes dupliquées, données manquantes...
- Des **confusions** peuvent avoir eu lieu **entre Session et Répétition**.
- Des **jeux de données atypiques** sont présents (jeux de données complètement déséquilibrés, avec une seule observation, avec un grand nombre de répétitions, ...)

Ces problèmes ont été traités et résolus. Cette étape de « nettoyage » de base de données est une étape essentielle avant toute méta-analyse.

B. META-ANALYSES

I. DEFINITION ET INTRODUCTION

Le terme méta-analyse est principalement utilisé en médecine, et son utilisation a récemment été étendue à l'analyse sensorielle. Les **méta-analyses regroupent les résultats de différentes études** dans le but d'identifier les potentielles sources de désaccord entre les résultats, des structures de résultats, ou des relations intéressantes qui pourraient apparaître en accumulant plusieurs études. La motivation essentielle d'une méta-analyse est **d'agréger** de l'information pour obtenir des **résultats plus puissants ou nouveaux**.

Le terme méta-analyse est le plus souvent associé aux **essais cliniques** ayant pour but d'étudier la relation de cause à effet entre un acte médical et l'évolution d'un état de santé chez des sujets humains. Dans ce cas, les études regroupées concernent **toutes le même traitement médical**, testé par plusieurs cliniques. Ce n'est pas le cas en analyse sensorielle, où les produits sont tous différents d'une étude à l'autre. Cependant, des éléments des méthodologies proposées (et les difficultés rencontrées) dans le cas de méta-analyses d'essais cliniques peuvent être **transposables en analyse sensorielle**.

L'objectif de cette partie est de **définir une méthodologie** des méta-analyses en analyse sensorielle en nous basant sur les **essais cliniques** d'une part **et les quelques méta-analyses effectuées dans le cadre de l'analyse sensorielle** d'autre part (Pineau (2006), Mammasse (2012), Brockhoff et al. (2015)). Cette méthodologie visera à être utilisée dans l'ensemble de cette thèse mais aura également vocation à **être généralisée** à d'autres bases de données que la SensoBase. La méthodologie obtenue sera ensuite **testée sur différents exemples** (nombre de zéros dans les données sensorielles, différences entre modèle fixe et modèle aléatoire de l'ANOVA).

II. HISTORIQUE DES META-ANALYSES

En 1904, Karl Pearson effectuait la toute première méta-analyse pour essayer de surmonter le problème de **puissance statistique** réduite dans les études d'échantillon de petite taille (Pearson, 1904). Cinquante ans plus tard, la première méta-analyse démontrant **un effet d'une intervention thérapeutique** fut publiée. Dans les années 1970, des techniques statistiques plus sophistiquées furent développées dans le domaine des **sciences sociales** (Egger & Smith, 1997) et le **psychologue** Glass introduisit le terme de méta-analyse pour la première fois en 1976. Depuis, le nombre de **méta-analyses d'essais thérapeutiques** a explosé : alors qu'entre 1976 et 1986 seulement 100 travaux de méta-analyses avaient été publiés, 431 méta-analyses ont été effectuées en 1991 (Cucherat, 1997).

Pineau (2006) fut le premier à adapter la méta-analyse au profil sensoriel. Ainsi, il utilisa une base de données sensorielle (version 2006 de SensoBase) pour étudier les performances des panélistes. Plus tard, Mammasse (2012) utilisa également la méta-analyse pour déterminer

le nombre de sujets et de répétitions nécessaires dans des études de profil. Enfin, Brockhoff et al. (2015) l'ont utilisé pour illustrer l'effet d'un nouveau modèle.

III. METHODOLOGIE POUR LES META-ANALYSES SUR ESSAIS CLINIQUES

Dans le cadre de méta-analyses d'essais cliniques, Nony et al. (1995) mirent en place une méthodologie recommandant de :

- **formuler l'objectif principal** de la méta-analyse : un seul objectif principal défini *a priori*.
- **définir les effets étudiés** (critères de jugement) : il est préférable que le critère de jugement soit un critère clinique, perceptible par le patient (diminution de la mortalité totale par exemple).
- **collecter les essais thérapeutiques disponibles** : cette étape est capitale et soumise à certains biais comme le biais de publication. En effet, les études qui ont obtenu les résultats espérés sont publiées plus souvent que les autres, les autres restent souvent confidentielles.
- **sélectionner des essais** selon trois critères : leurs caractéristiques, leur méthodologie et leur qualité.
- Demander d'éventuelles informations complémentaires.
- **analyser quantitativement** les résultats. Les techniques utilisées reposent sur l'hypothèse suivante : chaque essai réalise une estimation de l'effet réel du traitement à partir d'un échantillon de patients.
- **interpréter les résultats de la méta-analyse** : il faut rester particulièrement critique. La robustesse d'une méta-analyse doit être appréhendée par une analyse de sensibilité (utilisation de plusieurs techniques statistiques différentes, réalisation d'une méta-analyse cumulative, réalisation de plusieurs méta-analyses avec inclusion et exclusion de différents essais...).

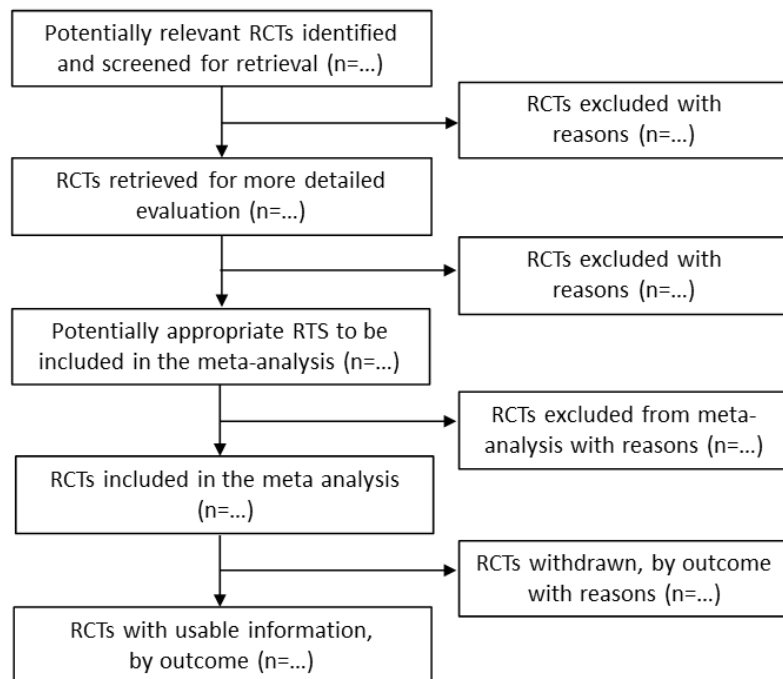
En 1999, une conférence sur la qualité des méta-analyses (Quality of Reporting of Meta-analyses, QUORUM) eut lieu (Moher et al., 1999) et déboucha sur une **check-list** pour publier une méta-analyse (extrait Figure 11 **Figure 1**), et sur un **diagramme de sélection des jeux de données** (Figure 12). Ce diagramme permet d'expliquer et de légitimer l'exclusion de certains jeux de données dans la méta-analyse. Ainsi, aucun jeu de données ne doit être exclu de l'analyse sans que la raison de son exclusion ne soit spécifiée et justifiée.

Heading	Subheading	Descriptor
Title		Identify the report as a meta-analysis (or systematical review) of RCTs*
Abstract		Use a structured format
	Objective	Describe the clinical question explicitly
	Data sources	The databases (list) and other information sources
	Review methods	The selection criteria (ie population, intervention, outcome and study design): methods for validity assessment, data abstraction and study characteristics, and quantitative data synthesis in sufficient detail to permit replication
	Results	Characteristics of the RCTs included and excluded: qualitative and quantitative findings (ie, point estimates and confidence intervals): and subgroup analysis
	Conclusion	The main results

Extract of The Lancet (1999)

*RCT : randomised control trials

FIGURE 11 : EXTRAIT DE LA CHECK-LIST POUR LA PUBLICATION D'UNE META ANALYSE (QUORUM)



The Lancet, 1999

FIGURE 12 : DIAGRAMME DE SELECTION DES JEUX DE DONNEES (QUORUM)

IV. META-ANALYSES DEJA EFFECTUEES SUR LES BASES DE DONNEES SENSORIELLES PREFBASE ET SENSOBASE

Les bases de données sensorielles PrefBase et SensoBase ont déjà été l'objet de méta-analyses lors des thèses de Pineau (2006) et Mammasse (2012).

1. META-ANALYSE DE PINEAU (2006)

Dans sa thèse, Nicolas Pineau propose une **démarche** pour effectuer ses méta-analyses :

- Définir une problématique
- Coder les données (unifier les informations, discrétiser certaines variables telles que l'âge...)
- Filtrer les données
- Pondérer les données (éventuellement)
- Choisir le modèle statistique à appliquer (en général, il s'agit du modèle linéaire)

Il utilise cette méthodologie pour étudier **la performance des sujets en analyse sensorielle** (discrimination, accord et répétabilité) (Pineau, 2006) et ses liens avec les caractéristiques des panélistes et des descripteurs. Ce thème sera également traité dans cette thèse. Selon cette étude, le niveau de discrimination s'élève avec le niveau scolaire, mais ce dernier ne modifie pas les niveaux d'accord et de répétabilité. L'âge agit significativement sur l'indice de discrimination uniquement, mettant en avant des performances plus élevées pour les personnes entre 20 et 45 ans. L'effet du type de descripteur est significatif pour tous les indices de performances. La répétabilité observée est meilleure pour les descripteurs de saveurs, d'arômes et d'odeur, mais ce résultat pourrait être lié à l'utilisation d'une portion plus restreinte de l'échelle de notation pour ces descripteurs. Les niveaux d'accord et de discrimination sont beaucoup plus élevés pour les descripteurs visuels. Le niveau d'accord est également meilleur pour les descripteurs de texture que pour les termes d'arôme de saveur et d'odeur, mais la discrimination de la texture est relativement faible. L'effet genre n'est jamais significatif. En revanche, les hommes auraient tendance à utiliser une portion de l'échelle plus importante que les femmes lors de la notation des descripteurs visuels.

Pour établir ces résultats, Pineau a effectué une ANOVA selon le modèle :

$$\text{performance} = \text{jeu de données} + \text{variable} + \text{variable} * \text{jeu de données}$$

avec le facteur « jeu de données » aléatoire. De plus, il a proposé une pondération des jeux de données (ou des sujets) originale par l'équilibre de représentation de tous les niveaux de la variable qualitative du modèle (genre par exemple) dans le jeu de données concerné. Cette pondération est le rapport de l'effectif minimum observé dans une catégorie et de l'effectif moyen par catégorie.

Pineau a également souligné certaines difficultés à analyser la SensoBase. En effet, il existe un grand nombre de données manquantes dans la SensoBase. Par ailleurs, comme expliqué précédemment, les jeux de données ne sont pas toujours équilibrés, et ce déséquilibre peut se révéler problématique lors des méta-analyses.

2. META-ANALYSE DE MAMMASSE (2012)

Mammasse (2012) a travaillé sur le nombre de sujets et le nombre de répétitions nécessaires en analyse sensorielle (sur les bases de données PrefBase et SensoBase). Elle a également utilisé la méthodologie définie par Pineau. Elle a étudié le nombre de sujets dans chacun des deux types de panel (profil et hédonique) par une approche de ré-échantillonnage qui consiste à réduire progressivement le nombre de sujets tant que les résultats de l'analyse statistique demeurent stables.

3. META-ANALYSE DE BROCKHOFF (2015)

Encore plus récemment, Brockhoff et al. (2015) ont utilisé 369 jeux de données de la SensoBase pour observer l'influence de l'utilisation du Mixed Assessor Model (MAM) (nouveau modèle qui sera décrit dans le Chapitre 3) plutôt que l'utilisation d'un modèle d'ANOVA à deux facteurs usuel. Pour cela, ils se sont basés essentiellement sur des techniques de comptage et de pourcentage (dénombrement du pourcentage de descripteurs à effet produit significatif avec le MAM et avec le modèle à deux facteurs pour différents seuils de significativité).

C. METHODOLOGIE POUR LES META-ANALYSES EN PROFIL SENSORIEL

I. PRESENTATION DE LA METHODOLOGIE ETABLIE

En regroupant les différentes méthodologies mises en place pour les essais cliniques et pour nos bases de données, certains éléments communs ont été mis en évidence et nous ont permis de construire notre propre méthodologie (Figure 13) vouée à être systématiquement appliquée à toute méta-analyse de cette thèse-.

Une première étape, commune à la méthodologie de Pineau et à celle des essais cliniques, consiste à définir la **problématique de l'étude**. Dans cette thèse, les problématiques seront essentiellement de comparer deux méthodes ou deux modèles suivant différents critères.

Une fois la problématique définie, nous pouvons réfléchir au **protocole** à mettre en place : autrement dit **quelles analyses effectuer** et **quels jeux de données sélectionner**.

Certaines analyses impliquent des conditions particulières (inversibilité de matrices, nombre de sujets ou de produits minimal, etc.). La non prise en compte de ces conditions donne lieu au mieux à l'arrêt immédiat du programme, au pire à des résultats incohérents qu'il sera difficile de détecter par la suite. Il est donc important de commencer par déterminer quelles analyses effectuer **AVANT** de sélectionner les jeux de données. En général, l'analyse consiste à parcourir l'ensemble des jeux de données et à calculer différents critères ou à mettre en place un modèle global.

La sélection des jeux de données doit prendre en compte **les conditions nécessaires aux critères** choisis préalablement. Dans un cas idéal, les conditions de sélection devraient être figées avant même de lancer l'analyse. Dans la pratique, cette sélection est souvent amenée à être révisée aux différentes étapes de la méta-analyse (apparition de jeux de données atypiques, cas particuliers impliquant l'impossibilité de calculer un critère statistique...).

Pour visualiser l'ensemble des jeux de données sélectionnés, nous conseillons de **caractériser** ensuite **la base de données obtenue**. En effet, les différences entre les différents jeux de données utilisés dans les méta-analyses doivent être décrites et maîtrisées au maximum. Ainsi, le nombre d'observations, de sujets, de produits, de répétitions ou d'attributs peuvent impacter directement l'analyse statistique d'un jeu de données. La distribution de ces paramètres permet de voir apparaître des jeux de données atypiques et/ou des groupes de jeux de données. Sinon, il est possible de faire des études en séparant les différents groupes de jeux de données. Certains jeux de données atypiques ou erreurs de saisie peuvent aussi être détectés. Cette caractérisation peut être effectuée avant et/ou après une mise en forme éventuelle de la base de données (pondération, homogénéisation, mise à la même échelle...).

La méta-analyse est ensuite conduite et des **résultats** sont obtenus. Si un critère est étudié au niveau du jeu de données, l'analyse de la distribution des critères donne des indications

claires sur la stabilité des résultats. Cette analyse permet également de repérer d'éventuels jeux de données atypiques et groupes de jeux de données aux résultats similaires. La mise en œuvre des calculs peut mettre également en évidence des difficultés. En effet, certains indicateurs peuvent parfois être difficilement calculables pour certains jeux de données (cas particuliers non prévus par les fonctions R implémentées). Ces jeux de données doivent être listés, puis analysés un par un pour évaluer si la fonction peut être adaptée pour prendre ce cas en compte ou si un critère de sélection doit être ajouté.

La stabilité du résultat peut être observée par la distribution des indicateurs (et notamment leur écart type). Par ailleurs, effectuer des études complémentaires avec d'autres méthodes statistiques est recommandé pour valider ses conclusions.

Enfin, il peut être intéressant de recommencer la méta-analyse sur un sous ensemble de jeux de données pour observer si on obtient bien les mêmes résultats.

Ces résultats sont ensuite **interprétés**. Si le comportement d'un des jeux de données est atypique, il est important de comprendre pourquoi il diffère des autres jeux de données avant, éventuellement, de recommencer l'analyse en excluant ce jeu de données.

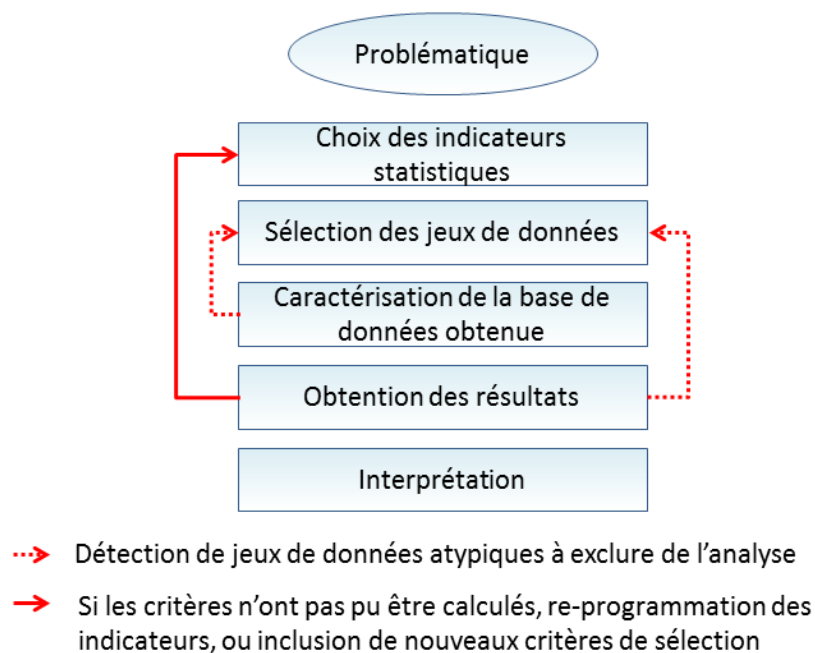


FIGURE 13 : METHODOLOGIE ADOPTÉE POUR LES META-ANALYSES DE CE MANUSCRIT

II. OUTILS DISPONIBLES POUR LA META-ANALYSE

Quel que soit l'objectif de la méta-analyse à effectuer, la caractérisation de la base de données sélectionnée est primordiale pour éviter de traiter des jeux de données atypiques et d'avoir des résultats incohérents. Par conséquent, différentes fonctions R ont été proposées afin de résumer les informations d'une base de données saisie en paramètre.

Le premier outil permet d'obtenir la liste des caractéristiques (nombre de sujets, nombre de produits, nombre de répétitions, nombre d'attributs, équilibré ou non par jeu de données).

Le second outil synthétise cette liste sous forme d'histogrammes. C'est ce second outil qui permet de visualiser les groupes de jeux de données ou les jeux de données atypiques.

Ces deux outils sont intégrés dans un package et prennent en paramètre une base de données dont les colonnes sont les codes jeux de données, sujets, produits, descripteurs, répétition, session et note. Ces deux fonctions renvoient des fichiers html directement importables dans Word.

D. EXEMPLE : UTILISATION DE LA METHODOLOGIE DES META-ANALYSES SUR SENSOBASE

I. ETUDE DES ZEROS DANS LA SENSOBASE

1. PROBLEMATIQUE

La Figure 7 a montré un pourcentage élevé de zéros dans les notes de la SensoBase. La problématique de cette étude est d'analyser pourquoi ce pourcentage est aussi élevé. Cela pose la question de l'uniformité de la répartition de ces zéros dans l'ensemble des jeux de données et par modalité sensorielle des descripteurs.

2. PROTOCOLE

i. CRITERE ETUDIE

Nous avons calculé pour chaque jeu de données :

- Le **pourcentage de zéros dans chaque jeu de données** (tous attributs confondus)
- Le **pourcentage de zéros pour chaque descripteur**
- Les **moyennes et médianes de pourcentages de zéros** pour les différents **types de descripteur**

ii. SELECTION DES JEUX DE DONNEES

Nous avons sélectionné **1 167 jeux de données** avec plus d'une observation.

3. CARACTERISATION DE LA BASE DE DONNEES OBTENUE

La caractérisation est la **même que celle de la SensoBase** faite précédemment. Le récapitulatif de la méta-analyse est présent dans la Figure 6.

4. RESULTATS

La Figure 14 montre qu'il existe des jeux de données et des descripteurs avec un fort pourcentage de zéros. 54 jeux de données ont plus de 50% de zéros dans l'analyse. Ces jeux de données peuvent être problématiques lors d'analyses basées sur les ANOVAs.

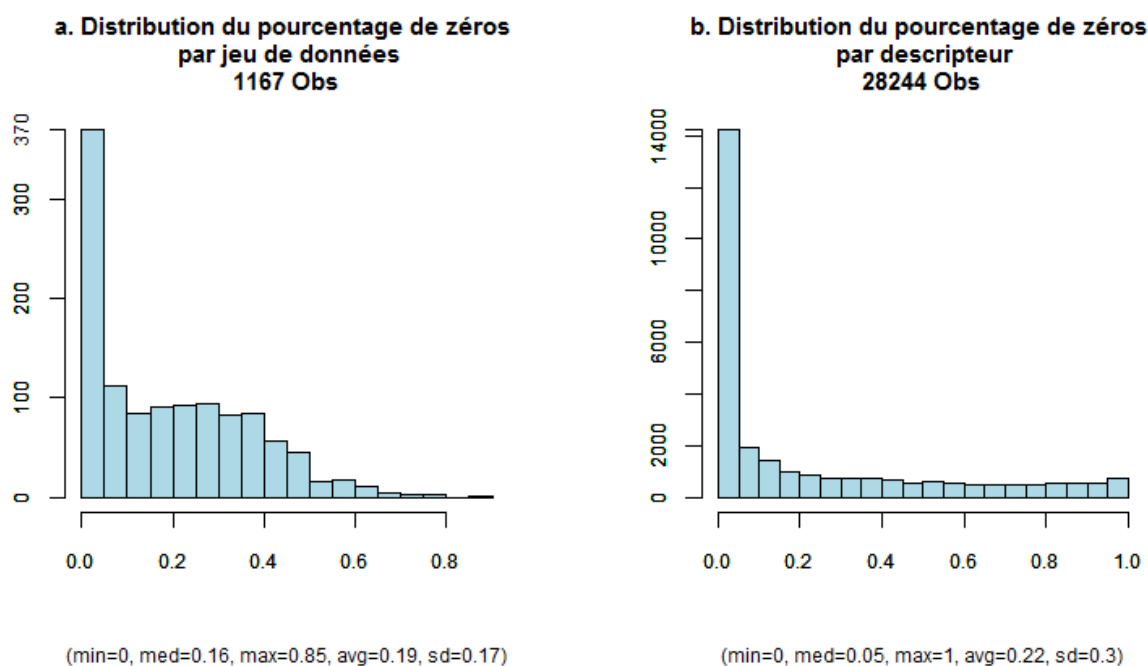


FIGURE 14 : DISTRIBUTION DU POURCENTAGE DE ZEROS PAR JEU DE DONNEES ET PAR DESCRIPTEUR

Par ailleurs, le nombre de zéros par descripteur semble varier en fonction de la modalité sensorielle du descripteur (Tableau 18).

Modalité sensorielle	Nombre de descripteurs	Moyenne	Médiane
Visuel	2 956	6 %	0 %
Odeur	4 974	28 %	16 %
Saveur	4 077	14 %	14 %
Texture	4 933	8 %	0 %
Aromes	9 407	34 %	23 %
Arrière-gout	1 017	22 %	8 %
Trigéminal	723	13 %	3 %
Hédonique	7	0 %	0 %
Autre	150	18	12 %

TABLEAU 18 : POURCENTAGES MOYENS ET MEDIANS DE ZEROS PAR MODALITE SENSORIELLE

Les résultats ci-dessus montrent que le pourcentage de zéros dans les notes attribuées est plus fort pour les descripteurs d'arômes et d'odeur.

5. COMMENTAIRES

Le nombre de zéros dans les données de profil sensoriel est assez impressionnant, et pourrait être lié au fait que certains descripteurs sont introduits alors qu'ils ne

correspondent pas (ou pas bien) à certains produits de l'étude. En effet, certains laboratoires utilisent systématiquement la même liste de descripteurs quelle que soit la sélection de produits étudiés. Pour les jeux de données à plus de 50 % de zéros, la moyenne du nombre de descripteurs est de 38 contre 23 descripteurs seulement pour les jeux de données à moins de 50 % de zéros.

II. ETUDE DE L'IMPACT DU TYPE D'EFFET SUJET (FIXE OU ALEATOIRE) DANS L'ANOVA

1. PROBLEMATIQUE

Comme présenté dans le Chapitre 1 (II), il est recommandé d'effectuer des ANOVA à 2 facteurs avec effet sujet aléatoire (en d'autres termes de comparer l'effet produit au désaccord entre les sujets). Cependant, ceci est basé sur le fait que **le désaccord entre les sujets doit être plus grand que la non-répétabilité entre les sujets**. En effet, les différences intra-individuelles (mesurées par les carrés moyens de l'erreur) sont censées être plus petites que les différences inter-individuelles (mesurées par les carrés moyens de l'interaction). En d'autres termes, pour un produit, un descripteur et une répétition fixée, il y a théoriquement moins de différences entre les deux notes d'un même sujet qu'entre les notes de différents sujets. Cette méta-analyse permet de vérifier si c'est bien le cas.

2. PROTOCOLE

i. CRITERE ASSOCIE

Les carrés moyens de l'erreur et de l'interaction CM_{erreur} et $CM_{interaction}$ sont calculés et comparés. Ainsi le **pourcentage de descripteurs** dont le CM_{erreur} est plus grand que le $CM_{interaction}$ est obtenu. Les **changements de diagnostic** de significativité seront également étudiés ($\alpha=0.05$).

ii. JEUX DE DONNEES SELECTIONNES

On sélectionne exclusivement les jeux de données avec **au moins deux répétitions et strictement plus de deux produits**. En effet, l'objectif de l'ANOVA est de comparer plus de deux moyennes entre elles. Dans le cas où l'on cherche à détecter des différences entre deux produits, on préfère généralement utiliser un test de Student (t-test).

3. CARACTERISATION DE LA BASE DE DONNEES

697 jeux de données ont été sélectionnés avec les caractéristiques présentées dans le Tableau 19, Tableau 8 et la Figure 15.

Nombre de sujets	Nombre de produits	Nombre de descripteurs
4 429	4 425	14 540

TABLEAU 19 : CARACTERISTIQUES DE LA BASE DE DONNEES OBTENUE

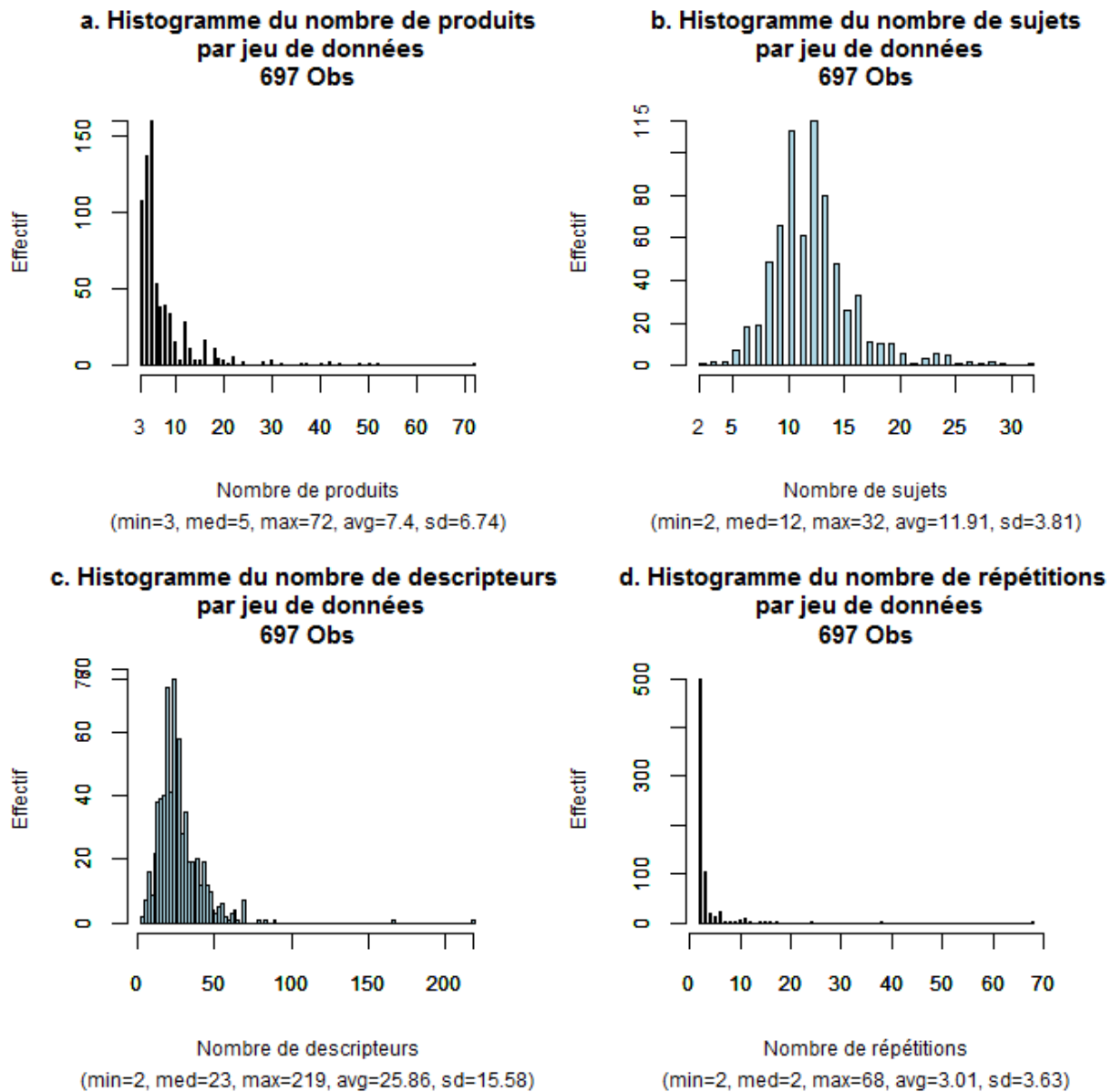


FIGURE 15 : CARACTERISTIQUES DES 697 JEUX DE DONNEES SELECTIONNES

4. RESULTATS ET INTERPRETATION

17 918 descripteurs ont été étudiés.

Pour **21.56 %** de ces descripteurs, on a observé un carré moyen de l'erreur **plus grand que le carré moyen de l'interaction**.

48.24 % des descripteurs ont montré un effet produit significatif quel que soit le modèle choisi pour $\alpha = 0.05$ et **27.98 %** ont montré un effet produit non significatif quel que soit le modèle choisi.

Pour **0.48 %** des descripteurs, l'effet produit a été significatif ($\alpha=0.05$) avec le modèle à effet sujet aléatoire alors qu'il ne l'était pas avec le modèle à effet fixe. Pour **23.40 %** des

descripteurs, l'effet produit n'est pas significatif avec l'effet sujet aléatoire alors qu'il l'était avec effet fixe : c'est le résultat important de cette méta-analyse.

Pour plus d'un descripteur sur cinq, l'effet produit aurait pu être testé contre la résiduelle. Mais cela n'entraîne un changement de diagnostic que pour à peine 1 % des descripteurs. En revanche, considérer l'effet sujet comme aléatoire permet de ne pas conclure à tort à des différences entre les produits. Pour un descripteur sur cinq, le modèle fixe conclut à des différences entre produits alors que le modèle mixte ne le permet pas.

E. CONCLUSION

La méthodologie des méta-analyses mise en place est bien adaptée aux bases de données sensorielles. Les points clés de cette méthodologie sont :

- une **sélection** des jeux de données entièrement **justifiée** : les jeux de données pour lesquels les critères à calculer ne fonctionnent pas doivent impérativement être analysés, et, soit un critère de sélection doit être ajouté, soit le calcul du critère statistique doit être adapté.
- une **caractérisation systématique** de la base de données sélectionnée pour visualiser d'éventuels groupes ou point aberrants.
- une **vérification** des résultats obtenus grâce à la distribution des critères étudiés.

Cette méthodologie a été utilisée pour effectuer l'ensemble des méta-analyses de cette thèse (cf. Annexe 2).

CHAPITRE 3 :

NOUVELLES METHODES D'ANALYSE STATISTIQUE DES DONNEES DE PROFIL SENSORIEL

Dans ce chapitre, de nouvelles méthodes statistiques portant sur le contrôle des performances des sujets et sur les cartographies des produits sont proposées pour résoudre les problèmes mis en évidence dans le Chapitre 1. Elles sont ensuite illustrées à l'aide du jeu de données « cheeses ». Ces méthodes ont été implémentées sous forme de bibliothèques R et sont disponibles sur le site www.timesens.com.

A. CONTROLE DES PERFORMANCES

I. PERFORMANCES UNIDIMENSIONNELLES

1. PERFORMANCES UNIDIMENSIONNELLES USUELLES

i. PERFORMANCES DU PANEL ET DES SUJETS

Avant d'effectuer une épreuve de mesure de profil sensoriel, l'animateur de panel doit savoir si son panel et ses sujets sont performants lorsqu'ils attribuent leurs notes aux produits. Il est ainsi amené à tester la **discrimination** (aptitude à détecter des différences sensorielles entre produits), l'**accord** (aptitude à être en consensus avec le panel) et la **répétabilité de son panel et de ses sujets** (aptitude à donner des notes très proches à un même produit présenté plusieurs fois).

ii. PRESENTATION DE LA METHODE CAP

La **méthode CAP** (Control of Assessor Performances) permet d'estimer les performances du panel et des sujets. Elle présente les résultats des performances unidimensionnelles (descripteur par descripteur) du panel et des sujets sous forme d'une table synthétique. Elle a été développée par Schlich (1997b), puis mise à disposition des animateurs de panel inscrits au programme SensoBase (www.sensobase.fr).

La table CAP se divise en **deux parties** : les premières colonnes présentent les **performances du panel** dans sa globalité, les colonnes suivantes présentent les **performances individuelles**. Chaque ligne correspond à un descripteur.

La **discrimination du panel** est mesurée par le **F produit** de l'**ANOVA à deux facteurs** (produit et sujet) avec effet sujet aléatoire : si l'effet produit est significatif, le panel est discriminant. L'**accord du panel** est évalué par le test de l'**interaction produit*sujet** : si l'interaction est significative, le panel est en désaccord. Enfin, la racine du **Carré Moyen de l'Erreur (CME)** donne une indication sur la **répétabilité du panel** : plus le CME est petit, plus le panel est répétable. Les CME peuvent être comparés d'un jeu de données à un autre, du moins si les échelles de mesure sont identiques. Notons que les descripteurs sont ordonnés dans la table CAP du plus discriminant au moins discriminant selon le F décroissant.

Les performances de chaque sujet sont ensuite évaluées pour chaque descripteur. La **discrimination** est évaluée par le **F du modèle individuel à un facteur (produit)**. Le niveau

d'accord est calculé à l'aide d'un **coefficient de corrélation** (de Kendall) entre ses notes et celles des autres sujets discriminants. Enfin, un dernier calcul évalue le niveau de **répétabilité du sujet** par rapport au groupe. Pour cela, les écarts types des répétitions sont calculés par sujet et par produit (un écart-type par couple sujet/produit est calculé), puis un **test de contrastes** permet de tester si chacun des écarts types d'un sujet est statistiquement plus grand que ceux du reste du panel.

Pour chaque descripteur, un rang est attribué à chaque sujet en fonction de sa discrimination (le sujet avec le meilleur F aura le rang 1, le deuxième meilleur F aura le rang 2, etc.). La moyenne des rangs par sujet donne le **F-rank** et permet d'ordonner les colonnes des sujets en fonction de leur discrimination décroissante sur l'ensemble des descripteurs.

Un code couleur a été instauré pour permettre à l'utilisateur de visualiser directement les résultats des tests. Pour le panel, une couleur verte dans la colonne F-prod (ou F-Disag) indique que ce panel a été discriminant (ou en accord). Une couleur verte pour un panéliste indique qu'il a été discriminant et en accord avec le panel, une couleur rouge indique qu'il a été discriminant, mais en désaccord. Enfin, une couleur jaune indique que le panéliste n'a pas été discriminant.

iii. EXEMPLE DE TABLE CAP APPLIQUEE AU JEU DE DONNEES « CHEESES »

	Mean F-Prod	F-Disag	RMSE	S5742	S5735	S5736	S5737	S5746	S5744	S5733	S5741	S5739	S5734	S5740	S5738	S5747	S5748	S5745	S5743	
Salted	3.67	53.56	2.77	1.26	+ lr	+	+	+	+	+	+	+	+	+	+	+	+	= lr	=	+
Hard	4.59	46.97	1.8	1.19	+	+	+	+	+	+	+	+	+	+	+	+	=	+ lr	=	=
Dense	5.63	27.61	3.24	1.15	+ lr	+	- lr	+	+	+	+	+	+	+	= lr	=	+ lr	+	=	=
Intense	3.61	27.37	2.49	1.17	+	+	+	+	+	+	+ lr	+ lr	=	+	=	+	= lr	=	=	=
Crumbly	5.97	25.22	2.65	1.26	+	+	+	+	= lr	+	+	=	=	+	= lr	+	=	= lr	+	=
Smooth	3.78	18.32	2.75	1.17	+	+	+	+	+	+	+	= lr	=	=	=	=	= lr	=	=	=
Elastic	5.42	16.45	2.6	1.46	=	+	+	+	= lr	+	+	=	+	+	=	=	= lr	+ lr	+	=
Fat	4.22	13.95	1.97	1.33	+ lr	+	- lr	= lr	+	=	= lr	+	+ lr	+	=	+	=	+ lr	= lr	=
Sticky	4.11	13.17	1.71	1.52	+	= lr	=	+	+	+	= lr	+	= lr	+	= lr	+	+	=	=	=
Butter	3.58	11.64	1.82	1.27	+	+	+	+	=	=	+	+	=	=	=	=	=	= lr	=	=
Fruity	2.71	9.77	2.35	1.15	+	+	+	=	+	=	+	+	= lr	= lr	= lr	=	+	= lr	=	=
Milky	3.73	7.14	1.09	1.33	+	+	+ lr	= lr	=	=	=	=	= lr	= lr	=	=	=	= lr	=	=
F Rank	-	-	-	-	3.08	3.58	3.83	5.92	7.08	7.75	7.83	8	9.33	9.58	9.58	10.5	11	11.08	13.25	14.58

Panel Performances		Assessor Performances
F-Prod	F-Disag	+ Discriminant and in agreement
p > 0.05	p < 0.05	- Discriminant and in disagreement
p < 0.05	p > 0.05	= Not Discriminant
Repeatability: lr=Less repeatable		

NA's shows a problem in calculations (no variability). 0 is returned when all products have the same scores.

FIGURE 16 : TABLE CAP DU JEU DE DONNEES CHEESES

La table CAP du jeu de données « cheeses » (Figure 16) montre que les produits sont discriminés par le panel pour tous les descripteurs, avec des F produits élevés attestant de fortes différences entre les produits. Le descripteur le plus discriminant est Salted et le moins discriminant est Milky. Le désaccord est significatif au niveau du panel pour tous les descripteurs sauf Milky mais faible en regard des F produits. Les racines des CME (RMSE)

attestent de la répétabilité moyenne du panel : sur une échelle de 0 à 10, il y a en moyenne 1.2 point d'écart entre 2 évaluations par un même sujet pour un même descripteur. Les deux descripteurs pour lesquels le panel a été le plus répétable sont Fruity et Dense (RMSE=1.15). Sticky est le descripteur pour lequel le panel a été le moins répétable (RMSE=1.52).

En ce qui concerne les sujets, le sujet S5742 a été le plus discriminant (plus petit F-Rank). Il a discriminé les produits sur tous les descripteurs sauf Elastic et a été systématiquement en accord avec le panel. Ce sujet est aussi répétable que les autres pour tous les descripteurs sauf Salted, Dense et Fat.

On peut également remarquer que, souvent, bien que le panel soit en désaccord (case rouge pour les performances du panel), il existe très peu de sujets en désaccord avec le panel (cases rouges dans les performances des panélistes). Cela est dû à la différence d'approche entre les deux méthodes statistiques utilisées. Une case rouge au niveau du panel indique que le désaccord entre les sujets est plus grand que leur non-répétabilité, alors qu'une case rouge au niveau d'un sujet indique une mauvaise corrélation entre les notes du sujet et celles du panel.

A partir de cette table CAP, l'animateur de ce panel pourrait conclure que le panel a été bien discriminant sur l'ensemble des descripteurs mais que les trois derniers juges ne semblent pas être aussi performants que les autres (particulièrement S5748 qui est moins répétable que les autres sujets pour presque tous les descripteurs).

En conclusion, la table CAP est un outil pratique pour l'animateur de panel durant la **phase d'entraînement de ses sujets** : elle permet de **visualiser l'ensemble des performances** du panel et du sujet en un coup d'œil. Dans le cadre de cette thèse, la table CAP a été programmée sous R, intégrée dans une librairie (MAMCAP) et est désormais disponible dans le logiciel TimeSens.

2. SCALING ET MIXED ASSESSOR MODEL

i. INTRODUCTION AU SCALING

Dans le modèle d'ANOVA à 2 facteurs (produit et sujet) utilisé pour calculer les performances du panel, l'effet sujet est corrigé, ce qui permet de prendre en compte le fait que certains sujets attribuent des notes plus élevées (ou plus basses) que d'autres. En revanche, **le fait qu'un sujet disperse plus ou moins ses notes sur l'échelle** (on parle d'**effet de scaling**) n'est pas pris en compte. La Figure 17 illustre cet effet : le sujet A disperse plus ses notes que le panel pour les produits P1, P2 et P3 et le sujet B les disperse moins.

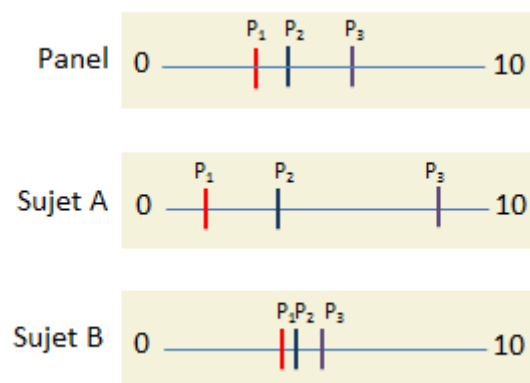


FIGURE 17 : EXEMPLES D'EFFETS DE SCALING

L'écart-type des notes étant un indicateur naturel de leur dispersion, on pourrait envisager de réduire (diviser par l'écart type) les données par sujet pour s'affranchir de l'effet de scaling. Ainsi, chaque panéliste aurait la même dispersion de ses notes dans l'analyse. Cependant, cette méthode ne prend pas en compte le désaccord éventuel du sujet avec le panel, ce qui constitue un désavantage illustré par la Figure 18.

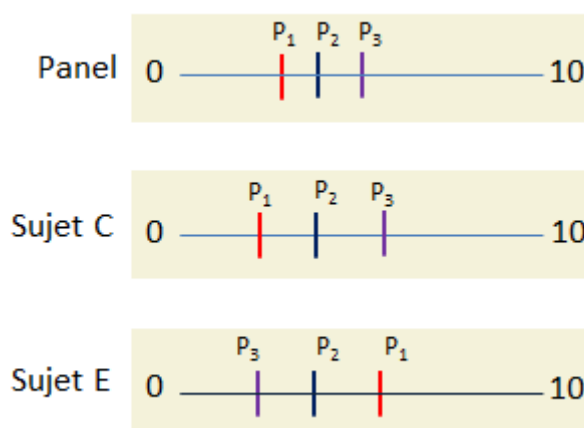


FIGURE 18 : DISPERSION DES NOTES ET ECARTS-TYPES

Dans ce cas, les sujets C et E ont le même écart type. Cependant, ***cela ne signifie pas qu'ils utilisent l'échelle de la même manière.*** En effet, le sujet E est en désaccord avec le sujet C et le panel. On peut penser alors que la dispersion des notes du sujet E ***n'est pas due à un effet d'échelle, mais à un désaccord avec le panel.*** Cette correction ne correspond donc pas à l'effet de scaling tel qu'il est perçu intuitivement, et l'écart-type ne peut par conséquent pas être utilisé comme indicateur de scaling.

En revanche, il est possible de considérer la dispersion des notes d'un sujet par rapport à celles du panel. Considérons le graphique dont les abscisses sont les notes du panel (centrées) et les ordonnées celles du sujet (centrées) (Figure 19) et plaçons les points correspondant aux notes obtenues pour chaque produit. La pente de la droite de régression

des points obtenus est appelée **coefficient de scaling**. Il peut être utilisé comme indicateur d'effet de scaling.

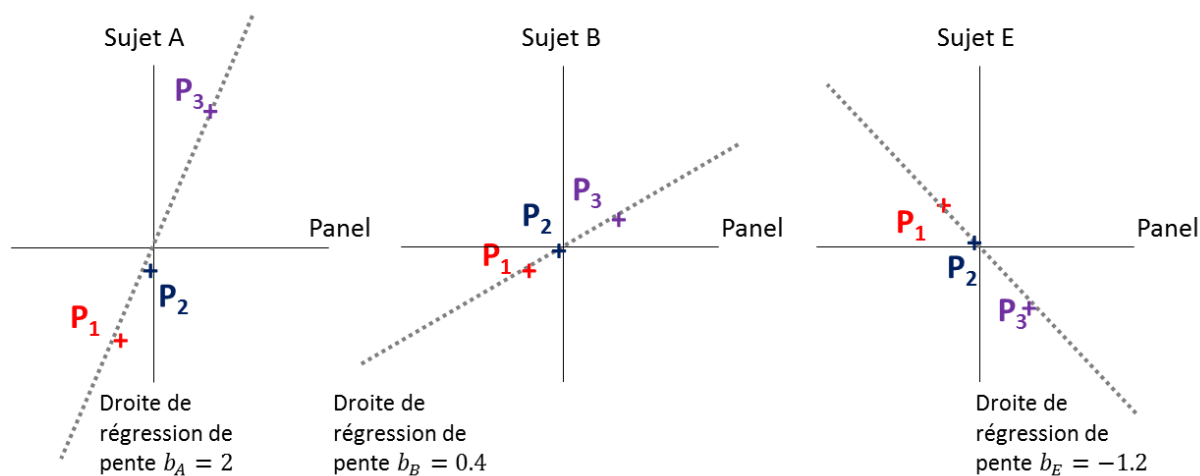


FIGURE 19 : EFFET DE SCALING ET COEFFICIENT DE REGRESSION

Lorsque le coefficient de scaling est plus élevé que 1 (cas du sujet A dont le coefficient de scaling est b_A), le sujet disperse plus ses notes que le panel. Lorsque le coefficient est compris entre 0 et 1 (cas du sujet B dont le coefficient de scaling est b_B) le sujet disperse moins ses notes que le panel.

Enfin, l'existence de coefficients de scaling négatifs (cas du sujet E) peut s'expliquer par plusieurs raisons :

- le sujet est en réel désaccord avec le panel : les produits ont été notés dans le sens inverse du panel ! Cela peut arriver si le sujet a mal compris les libellés des échelles, ou la signification du descripteur, ou s'il est affecté d'une perte de sensibilité relative au descripteur concerné.
- Si le panel et/ou le sujet n'ont pas été discriminants, alors l'ordre des produits sur les axes n'est pas forcément le même, et cela peut mener à des pentes négatives.

Un coefficient de scaling négatif ne doit par conséquent pas être interprété comme un effet d'échelle.

3. MIXED ASSESSOR MODEL (MAM)

Brockhoff et al. (2015) ont introduit ce coefficient de scaling dans un modèle statistique en 2015 : le **Mixed Assessor Model (MAM)**. Le MAM consiste à décomposer y_{ijk} (note attribuée par le sujet i au produit j à la répétition k) de la manière suivante :

$$y_{ijk} = \mu + \alpha_i + \gamma_j + b_i x_j + d_{ij} + \varepsilon_{ijk}$$

avec, μ la moyenne, α_i l'effet sujet, γ_j l'effet produit, b_i le coefficient de scaling, $x_j = y_{.j} - y_{...}$, d_{ij} le désaccord pur, et ε_{ijk} le terme d'erreur.

Ce modèle présente la particularité de **décomposer l'interaction** usuelle de l'ANOVA à deux facteurs produit*sujet **en la somme d'un terme d'effet de scaling $b_i x_j$ et un terme d'effet de désaccord pur d_{ij}** . L'utilisation de la covariable x_j (notes centrées du panel) permet d'exprimer les notes du sujet en fonction des notes du panel comme présenté Figure 19.

Dans le cas d'un coefficient de scaling négatif (ne correspondant pas à un effet de scaling), l'interaction ne sera pas décomposée.

4. TESTS DE PERFORMANCES PRENANT LE SCALING EN COMPTE

De nouveaux indicateurs ont été mis en place afin d'évaluer les performances du panel et des sujets **en prenant l'effet de scaling en compte**. Ces indicateurs sont basés sur des calculs de p-value de F (d'ANOVA) issus du modèle MAM (Tableau 20)

	Panel	Sujet
Discrimination	$CM_{product}/CM_{disag}$	$CM_{product}^i/CM_{error}^i$
Scaling	$CM_{scaling}/CM_{disag}$	$CM_{scaling}^i/CM_{disag}^i$
Accord	$CM_{desaccord}/CM_{error}$	$CM_{desaccord}^i/CM_{error}^i$
Répétabilité	-	$CM_{error}^i/CM_{error}^{(i)}$

TABLEAU 20 : F UTILISES POUR LES TESTS DE PERFORMANCES BASES SUR LE MAM

$CM_{product}$, CM_{disag} , $CM_{scaling}$, CM_{error} sont respectivement les carrés moyens associés aux effets produit, désaccord pur, scaling et erreur dans le modèle au niveau du panel.

$CM_{product}^i$, $CM_{scaling}^i$, CM_{disag}^i , CM_{error}^i sont respectivement les carrés moyens associés aux effets produit, scaling, désaccord et erreur **restreints au sujet i** et

$$CM_{error}^{(i)} = \frac{\sum_{i \neq i}^{I,J,K} (y_{ij.} - y_{ijk})^2}{(I-1)J(K-1)}$$

est le carré moyen associé à l'erreur dans le reste du panel (ne contenant pas le sujet i).

On peut noter que $CM_{scaling} = \frac{SC_{scaling}}{I-1}$ et $CM_{desaccord} = \frac{SC_{desaccord}}{(J-2)(I-1)}$, car on perd $(I-1)$ degrés de liberté dans le calcul du désaccord pur. $SC_{scaling}$ et $SC_{desaccord}$ sont les sommes des carrés associé aux effets de scaling, et de désaccord. En effet, ces degrés de liberté ont été utilisés pour calculer les I coefficients de scaling.

5. LA TABLE MAM-CAP

La table **MAM-CAP** (Peltier et al., 2014) s'inspire de la table CAP pour visualiser les résultats des performances obtenus par les tests associés au MAM. Le même code couleur est utilisé.

Ainsi, pour le panel, une couleur verte (ou rouge) indique une bonne (ou mauvaise) performance. La colonne F-Scal indique la présence ou non d'hétérogénéité de Scaling. Pour les sujets, une couleur jaune indique un sujet non discriminant alors qu'une couleur verte (ou rouge) indique un sujet discriminant et en accord (ou désaccord) avec le panel. Les tirets à l'intérieur des cases donnent des indications sur la force de discrimination des produits par les panélistes. Enfin, les crochets à l'intérieur des cases des sujets indiquent la présence d'un effet de scaling significatif (<> si le sujet a tendance à sur-disperser et > < s'il a tendance à sous-disperser ses notes).

Les résultats de la table MAM-CAP sur le jeu de données « cheeses » sont présentés Figure 20.

	Mean	F-Prod	F-Scal	F-Disag	RMSE	S5742	S5735	S5736	S5737	S5746	S5744	S5733	S5741	S5739	S5734	S5740	S5738	S5747	S5748	S5745	S5743	
Salted	3.67	216.71	16.23	0.68	1.26	<>	<	<	<	<	<	<	<	<	<	<	<	<	<	<	<	<
Hard	4.59	93.86	5.99	0.9	1.19																	
Intense	3.61	52.92	5.67	1.29	1.17																	
Crumbly	5.97	52.5	6.79	1.27	1.26																	
Dense	5.63	48.65	5.08	1.84	1.15																	
Smooth	3.78	47.13	8.86	1.07	1.17																	
Elastic	5.42	30.53	5.58	1.4	1.46																	
Fat	4.22	18.4	2.71	1.49	1.33																	
Butter	3.58	15.53	2.79	1.36	1.27																	
Sticky	4.11	14.07	1.36	1.6	1.52																	
Fruity	2.71	11.86	2.07	1.94	1.15																	
Milky	3.73	8.82	2.35	0.88	1.33																	
F Rank	-	-	-	-	-	3.08	3.58	3.83	5.92	7.08	7.75	7.83	8	9.33	9.58	9.58	10.5	11	11.08	13.25	14.58	

Panel Performances			Assessor Performances
F-Prod	F-Scal	F-Disag	Product discrimination: --- p< 0.01 -- p<0.05 - p<0.1
p> 0.05	p< 0.05	p< 0.05	Scaling Information: Average scale; >> Significant Smaller Scale;
p< 0.05	p> 0.05	p> 0.05	<< Significant Larger Scale; ? ? Negative scale
			Agreement:
			Yes No No Test
			Repeatability: lr=Less repeatable

FIGURE 20 : TABLE MAM-CAP DU JEU DE DONNEES CHEESES

L'analyse de cette table MAM-CAP montre que les produits ont été discriminés sur l'ensemble des descripteurs par le panel. Il existe une hétérogénéité de scaling pour tous les descripteurs sauf Sticky et le désaccord a été significatif pour Dense, Elastic, Fat, Sticky et Fruity. Le sujet S5742 a été le plus discriminant (plus petit F-Rank). Il a discriminé tous les descripteurs sauf Elastic. Pour le descripteur Salted, il a utilisé une échelle significativement plus grande que le panel. De plus, il a été de nombreuses fois en désaccord avec le panel. Il a également été moins répétable que le panel pour les descripteurs Salted, Dense et Fat.

On observe que certains sujets utilisent une échelle plus large pour la plupart des descripteurs (S5735), ou plus petite pour la plupart des descripteurs (S5747). En revanche, pour d'autres sujets l'échelle utilisée sera plus large pour certains descripteurs, et moins large pour d'autres (S5739).

Des différences entre les résultats obtenus par la table MAM-CAP et la table CAP doivent être soulignées. Tout d'abord, l'ordre des descripteurs change par rapport à la table CAP. Par exemple, on peut remarquer que Intense est le troisième descripteur plus discriminant alors qu'il était le cinquième dans la table CAP. Ainsi, **retirer l'effet de scaling a eu plus d'effet pour certains descripteurs** (Intense par exemple) que pour d'autres. Ensuite, les **F produit sont systématiquement beaucoup plus élevés dans la table MAM-CAP** que dans la table CAP : Salted est le plus discriminé (F à 216 pour la table MAM-CAP contre 53.56 pour la table CAP) et Milky a été le moins discriminé (8.82 pour la table MAM-CAP contre 7.14 pour la table CAP). Enfin, on observe **7 descripteurs avec un effet de désaccord significatif pour le modèle à deux facteurs et non significatif pour le modèle MAM**. Pour ces descripteurs, le désaccord était donc principalement dû à des effets de scaling de certains panélistes. La colonne RMSE de la table MAM-CAP et de la table CAP sont identiques.

Les tests de discrimination individuelle utilisés dans la table CAP et dans la MAM-CAP sont également identiques. En revanche, les **tests de désaccord sont différents** et on obtient certains résultats contraires à ce qui est obtenu dans la table CAP (désaccord du sujet S5742 sur plusieurs descripteurs). Pour ces descripteurs, le désaccord du sujet avec le panel est significativement plus élevé que son « désaccord avec lui-même » (répétabilité), mais les corrélations globales entre ses notes et celles du panel restent bonnes.

La table MAM-CAP est disponible via la **librairie R MAMCAP** accessible sur le site www.timesens.com. Cette librairie permet de calculer la table MAM-CAP et la table CAP. Il était en effet préférable de laisser la table CAP à disposition d'éventuels utilisateurs ne souhaitant pas utiliser le MAM. De plus, pour évaluer les performances d'un panel ayant dégusté seulement deux produits, la table MAM-CAP ne fonctionne pas car au moins 3 produits sont nécessaires pour calculer une droite de régression non triviale. Dans ce cas, nous conseillons d'effectuer une table CAP. Le test de corrélation ne pourra pas être effectué, mais la couleur verte (ou rouge) sera attribuée si le coefficient de corrélation associé est de 1 (ou -1).

6. DISCUSSION

Plusieurs limites inhérentes au modèle MAM ont été identifiées. La principale concerne l'accord entre les sujets. En effet, les tests présentés dans Peltier et al. (2014) permettent de tester si un sujet est « plus en désaccord » qu'il n'est répétable. Cela signifie qu'un sujet extrêmement répétable sera souvent en désaccord avec le panel. Pour pallier ce problème, nous avons introduit **une nouvelle option qui permet de tester l'accord du sujet avec le**

panel à l'aide de corrélations, comme dans l'approche CAP. Depuis la parution de l'article (Peltier et al., 2014), la librairie MAMCAP a intégré cette nouvelle fonctionnalité.

Si le coefficient de scaling obtenu correspond à un biais d'utilisation de l'échelle du sujet, la question du lien entre dispersion des notes et répétabilité se pose : l'effet d'échelle devrait également s'appliquer sur les répétitions (et donc la répétabilité). Ainsi, un sujet pourrait être moins répétable que le reste du panel à cause d'un effet d'échelle, simplement parce qu'il a tendance à disperser plus l'ensemble de ses notes. La question de la correction de l'effet de scaling dans la répétabilité se pose donc. Ainsi, une répétabilité absolue, constituée des CM_{error}^i/b_i pourrait être calculée. Cette quantité représenterait alors une **répétabilité du sujet indépendamment de son éventuelle tendance à sur ou sous disperser ses notes**. Une autre solution envisageable pour prendre en compte les répétitions dans le calcul du coefficient de scaling serait de calculer le coefficient de scaling à partir d'un nuage de points représentant les couples (produit, répétition). Ces pistes de réflexions pourront être abordées dans des travaux ultérieurs.

Enfin, le MAM présuppose que toute hyper- ou hypo-dispersion des notes est due à un effet d'utilisation d'échelle. Ce modèle offre donc **une correction systématique du scaling**. L'exemple cheeses montre que certains sujets semblent effectivement avoir une tendance à sur-disperser ou sous-disperser leurs notes quel que soit le descripteur. En revanche, pour d'autres sujets, **la tendance de disperser plus ou moins ses notes semble varier suivant les descripteurs**. Dans ce cas, peut-on légitimement parler d'effet d'échelle ? La forte dispersion d'un sujet sur l'échelle est-elle alors due à un effet d'échelle ou plutôt à un désaccord avec le panel ? Afin de répondre à ces questions, une alternative au scaling du MAM est proposée dans la partie suivante (II), en considérant le scaling comme un effet multidimensionnel.

II. LES PERFORMANCES MULTIDIMENSIONNELLES

1. LE MULTIMAM

Jusqu'ici, le MAM a été appliqué descripteur par descripteur. Nous proposons **d'étendre ce modèle à l'ensemble des descripteurs simultanément** à l'aide d'un nouveau modèle : le **MultiMAM** (Peltier et al., 2015e).

Si on note $Y_{ijk} = (y_{ijk}^d)$ le vecteur à d lignes représentant l'observation correspondant au sujet i , au produit j et à la répétition k , l'extension naturelle du modèle unidimensionnel est :

$$Y_{ijk} = M + A_i + P_j + \beta_i X_j + D_{ij} + E_{ijk}$$

Avec $M = \begin{pmatrix} \mu^1 \\ \vdots \\ \mu^D \end{pmatrix}$ la moyenne, $A_i = \begin{pmatrix} \alpha_i^1 \\ \vdots \\ \alpha_i^D \end{pmatrix}$ l'effet sujet, $P_j = \begin{pmatrix} \gamma_j^1 \\ \vdots \\ \gamma_j^D \end{pmatrix}$ l'effet produit,

$$X_j = Y_{j.} - Y_{...} = \begin{pmatrix} y_{j.} - y_{...}^1 \\ \vdots \\ y_{j.} - y_{...}^D \end{pmatrix}, D_{ij} \text{ le désaccord pur, et } E_{ijk} = \begin{pmatrix} \varepsilon_{ijk}^1 \\ \vdots \\ \varepsilon_{ijk}^D \end{pmatrix} \text{ le terme d'erreur.}$$

et β_i le coefficient de scaling (expliqué ci-dessous).

Le calcul du coefficient de scaling β_i peut alors se faire par deux approches :

- En considérant un coefficient de scaling par descripteur*subject (approche MAM) : β_i est alors un vecteur constitué des coefficients de scaling obtenu unidimensionnellement. On le nommera **coefficient de scaling multivarié** et le modèle associé à ce choix sera nommé « **MultimAM multivarié** ».
- En considérant un coefficient de scaling par sujet, tous descripteurs confondus. β_i est alors un scalaire que l'on nommera **coefficient de scaling global** : le modèle associé à ce choix sera nommé « **MultimAM global** ». Ainsi, la tendance d'un sujet à disperser les notes sera globale, valable pour l'ensemble des descripteurs.

Le **calcul du coefficient de scaling global** se fait exactement de la même manière que celui du coefficient de scaling unidimensionnel, mais tous descripteurs confondus. En d'autres termes, il s'agit de tracer en abscisses les notes centrées (par descripteur) du panel, en ordonnées celles du sujet pour l'ensemble des descripteurs, puis de placer les points correspondants aux produits, et enfin de tracer la droite de régression, comme indiqué Figure 21.

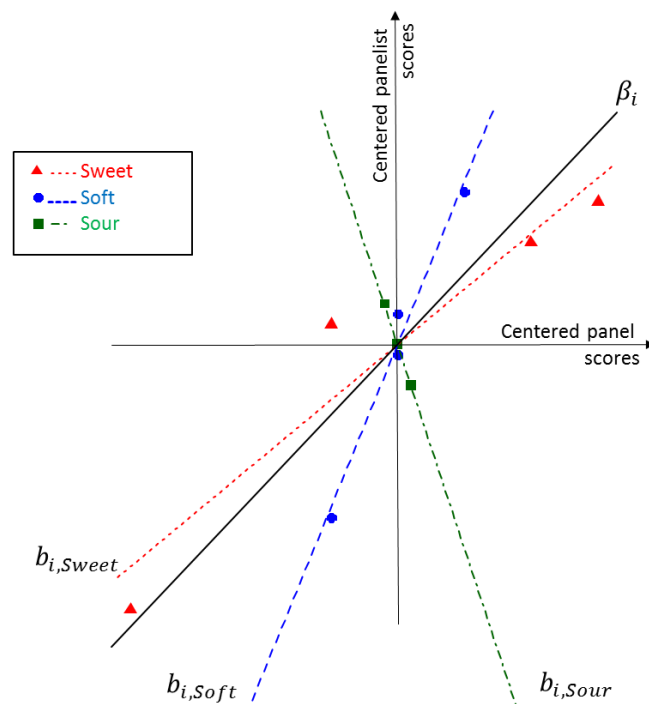


FIGURE 21 : OBTENTION DU COEFFICIENT DE SCALING GLOBAL POUR LE SUJET I

La **pente de la droite de cette régression** est appelée « **coefficient de scaling global** » ou « **overall scaling coefficient** ». Comme le coefficient de scaling univarié (MAM), un coefficient de scaling global supérieur à 1 (ou entre 0 et 1) indique que le sujet a tendance à sur-disperser (ou sous-disperser) ses notes par rapport au panel.

Un avantage de cette méthode réside dans le fait que **la régression se base sur plus de points que celle restreinte à un seul descripteur** (I*D points contre I points seulement). Avec cette méthode, si un descripteur est mal discriminé, l'ensemble des points correspondant à ce descripteur sera proche de zéro, et le descripteur en question aura peu d'influence dans la détermination de la pente de la droite (par exemple « Sour » sur la Figure 21). On observera théoriquement donc **moins de cas de négativité** de coefficients de scaling global avec cette méthode. En revanche un descripteur sur lequel de grandes différences ont été observées entre les produits aura une importance plus grande pour déterminer si un sujet note ou non plus large que le reste du panel.

L'exemple « cheese » est présenté dans Peltier et al. (2015e). Le coefficient de scaling global vaut 2.04 pour S5735 (qui sur-dispersait le plus souvent ses notes dans la table MAMCAP), et 0.42 pour S5747 (qui sous-dispersait le plus souvent ses notes dans la table MAMCAP).

2. TESTS DE PERFORMANCES MULTIDIMENSIONNELS

La MANOVA est **l'extension multidimensionnelle de l'ANOVA**. Pour obtenir les résultats d'une MANOVA, la matrice de discrimination $W^{-1}B$ est calculée, B étant la matrice de covariance de l'effet testé et W la matrice de covariance de l'erreur. Cette matrice est ensuite diagonalisée (calcul des valeurs et des vecteurs propres). Plusieurs statistiques peuvent être obtenues à partir des valeurs propres (statistiques de Hotelling-Lawley, Wilks, Pillai ou Roy). Ces statistiques sont ensuite ramenées au calcul d'un F suivant une loi de Fisher sous l'hypothèse de non-discrimination multidimensionnelle. Ces calculs permettent d'obtenir une p-value à partir des degrés de libertés de la matrice W .

Afin d'obtenir des tests de performances multidimensionnels, l'idée initiale était de remplacer les tests unidimensionnels basés sur l'ANOVA (Tableau 20) par des tests multidimensionnels basés sur la MANOVA.

En ce qui concerne l'extension des tests de performances des sujets, le nombre de degrés de liberté est souvent trop faible pour effectuer des MANOVA (nombre de descripteurs supérieur au nombre d'observations par sujet). La MANOVA n'est donc pas directement adaptée pour les performances multidimensionnelles individuelles. Une solution pourrait être de sélectionner uniquement certaines variables par sujet de manière à avoir moins de variables que d'observations. Cette approche était proposée sur le site www.sensobase.fr.

En ce qui concerne les tests du panel, l'adaptation pose également un certain nombre de problèmes.

Le **test de discrimination multidimensionnelle des produits à l'aide du modèle à deux facteurs** consiste à tester la matrice de covariance des produits contre la matrice de covariance de l'interaction à $(I - 1)(J - 1)$ degrés de liberté. Pour adapter ce test au MultiMAM, une première idée est de comparer la matrice de covariance de l'effet produit à la matrice de covariance du désaccord pur avec $(I - 1)(J - 2)$ degrés de libertés (extension naturelle de l'approche unidimensionnelle). On nommera cette méthode « **test de discrimination multidimensionnelle des produits** » globale ou **multivariée** selon l'option choisie dans le MultiMAM.

On constate que, lors du calcul du coefficient de scaling, $(I - 1)$ degrés de libertés associés au terme de désaccord sont perdus. Lors des calculs de la MANOVA, cette perte de degrés de liberté peut conduire à des tests beaucoup moins significatifs. Ainsi, le paradoxe suivant peut être observé : si on note B la matrice de covariance de l'effet produit, W la matrice de covariance de l'interaction à $(I - 1)(J - 1)$ degrés de liberté et D la matrice de covariance du désaccord à $(I - 1)(J - 2)$ degrés de liberté, le test comparant B et W peut être plus puissant que le test comparant B à D alors que le désaccord pur est systématiquement plus petit que l'interaction. La fréquence de ce paradoxe sera étudiée dans le chapitre 4.

Une solution proposée a été de corriger l'effet de scaling uniquement lorsqu'il est significatif. Considérons h_{ij}^d le terme de désaccord associé à cette solution (terme de désaccord pur si l'effet de scaling est significatif pour le descripteur d , terme d'interaction sinon), et supposons qu'une proportion p_{sig} des descripteurs soient significatifs pour l'effet de scaling. Ainsi, on note la matrice de covariance de désaccord $\tilde{W} = \tilde{X}'\tilde{X}$ où $\tilde{X} = \begin{pmatrix} h_{111}^1 & \cdots & h_{111}^D \\ \vdots & \ddots & \vdots \\ h_{IJK}^1 & \cdots & h_{IJK}^D \end{pmatrix}$. Les degrés de liberté associés à cette matrice \tilde{W} sont :

$$p_{sig}(I - 1)(J - 2) + (1 - p_{sig})(I - 1)(J - 1)$$

On note cette correction de scaling « **scaling conditionnel** », mais le fait qu'elle soit basée sur un calcul préalable de calcul de significativité de l'effet de scaling est discutable car cela consiste à utiliser les mêmes données plusieurs fois.

Par conséquent, notre tentative d'établir des tests de performance multidimensionnels basés sur la MANOVA n'a pas réellement abouti. Elle nous a cependant permis de mettre en évidence certains paradoxes de la MANOVA, principalement sur sa sensibilité au changement de degrés de liberté. L'exemple « cheese » n'illustre pas ce paradoxe, puisque les F de la MANOVA sont significatifs avec une p-value inférieure à 0.000 1 quel que soit le test de performance choisi (deux facteurs, multivarié, global ou corrigé).

3. UTILISATION DE LA MULTI-DIMENSIONNALITE POUR DETERMINER LES CAUSES DU SCALING

Lors de l'analyse de la table MAM-CAP de l'exemple « cheeses », on a pu observer que l'un des sujets (S5735) dispersait plus ses notes que le reste du panel. On peut supposer que si un sujet a tendance à disperser beaucoup ses notes sur l'échelle pour l'ensemble des descripteurs, cette tendance est plutôt **psychologique**. En revanche, si un sujet a tendance à disperser beaucoup (ou peu) ses notes pour un ou très peu de descripteurs seulement, on peut supposer que ce sujet est **hypersensible (ou hyposensible)** à ces descripteurs en particulier.

Ainsi, le scaling pourrait être **décomposé en une composante psychologique** commune à l'ensemble des descripteurs, **et une composante physiologique** spécifique à chaque descripteur. Cela consisterait à décomposer le coefficient de scaling MAM $\beta_{i,d}$ en un produit d'un coefficient de **scaling global** β_i (le même que celui présenté dans Peltier et al. (2015e) et d'un coefficient de **scaling corrigé** d'un éventuel biais psychologique du sujet (Peltier et al., 2015d).

$$\beta_{i,d} = \beta_i * \gamma_{i,d}$$

Le coefficient de scaling corrigé $\gamma_{i,d}$ peut permettre de **détecter des hyper- ou hyposensibilités qui ne sont pas détectées avec le coefficient de scaling MAM**. Supposons par exemple qu'un sujet disperse trois fois plus ses notes que le panel tous descripteurs confondus ($\beta_i = 3$). Supposons également que pour l'amertume, il disperse ses notes exactement de la même manière que le panel ($\beta_{i,amertume} = 1$). Alors le coefficient de scaling classique ne permet pas de détecter un effet de scaling pour ce descripteur. En revanche, le coefficient de scaling corrigé vaut 1/3 et montre une hyposensibilité du sujet pour l'amertume.

4. DISCUSSION

L'avantage du MAM est qu'il améliore systématiquement la discrimination des produits. Cependant, il peut arriver qu'une « sur ou sous-dispersion » ne soit pas due à un effet d'échelle mais à un réel désaccord du panéliste avec le panel. Dans ce cas, le MAM corrige les notes à tort. Le MultiMAM-global permet de réduire les risques d'une telle sur-corrrection en corrigeant uniquement l'effet de scaling commun à l'ensemble des descripteurs.

Afin de calculer un coefficient de scaling tous descripteurs confondus, il aurait pu être envisageable de considérer la moyenne des coefficients de scaling MAM. Cependant, la moyenne est sensible aux valeurs extrêmes, et encore plus aux coefficients de scaling négatifs. Le désavantage de cette méthode est donc qu'un calcul de la moyenne des coefficients de scaling MAM donne le même poids à chaque descripteur, même si certains

ne permettent pas de discriminer les produits. Le calcul du scaling global résout ce problème.

Le niveau de notation des sujets, mesuré par l'effet sujet en ANOVA, pourrait également être décomposé en une partie unidimensionnelle et une partie multidimensionnelle. Cette décomposition ne serait plus multiplicative mais additive :

$$\mu_{i,d} = \mu_i + \nu_{i,d}$$

μ_i serait alors le niveau de notation des notes global et $\nu_{i,d}$ le niveau des notes corrigé. On pourrait alors savoir si un sujet a tendance à noter haut tous descripteurs confondus à l'aide de μ_i , ou si cette tendance est spécifique à un descripteur pour lequel il ressentirait une intensité fixée plus forte que le panel. Le niveau de notes corrigé donne une information différente de celle apportée par le scaling corrigé. En effet, un fort niveau de notes corrigé indique qu'un sujet est très sensible à l'intensité de ce descripteur, alors qu'un fort scaling corrigé indique qu'un sujet est très sensible aux différences entre produits pour ce descripteur.

III. BILAN DES APPORTS SUR LES PERFORMANCES DU PANEL ET DES SUJETS

Afin de corriger l'effet de scaling, **de nouveaux tests de performances unidimensionnelles** basés sur le Mixed Assessor Model ont été proposés. Ces tests de performances ont été intégrés dans **une table synthétique** : la table MAM-CAP.

Une extension modèle multidimensionnelle du MAM a également été proposée (**MultiMAM**). Ce modèle propose deux options :

- (i) considérer un vecteur de coefficients de scaling par sujet en supposant que l'effet de scaling est dépendant des descripteurs (option **multivariée**),
- (ii) considérer un seul scalaire supposant l'effet de scaling indépendant des descripteurs (option **global**).

La mise en place de ce modèle n'a pas suffi à instaurer des tests de performances multidimensionnels satisfaisants. En revanche, elle a permis de mettre en évidence la **difficulté à évaluer les performances multidimensionnelles** à l'aide de la MANOVA. En effet, la **sensibilité de la MANOVA** au nombre de degrés de liberté du terme d'erreur ne permet pas de transposer simplement les tests unidimensionnels au multidimensionnel. Une solution alternative visant à ne corriger l'effet de scaling que lorsqu'il est significatif a été proposée.

L'utilisation du MultiMAM avec l'option « global » permet de calculer un nouveau coefficient de scaling. Il présente l'avantage d'être **unique par sujet** (tous descripteurs confondus) et de ne pas **sur-corriger** l'effet de scaling. Une méta-analyse étudiera dans quelles mesures les cas problématiques de négativité du coefficient de scaling MAM peuvent être résolus avec le MultiMAM.

Le coefficient de scaling MAM a également été décomposé en une composante reflétant **un biais psychologique** et l'autre **un biais physiologique**. Cette décomposition est en adéquation avec les mécanismes sensoriels du goût présentés dans le chapitre 1. Elle peut permettre à l'animateur de mieux connaître les caractéristiques de ses sujets. Ce principe peut également s'étendre au niveau des notes (effet sujet).

Finalement, on répertorie différents coefficients de scaling qui permettent de détecter différents effets (Figure 22).

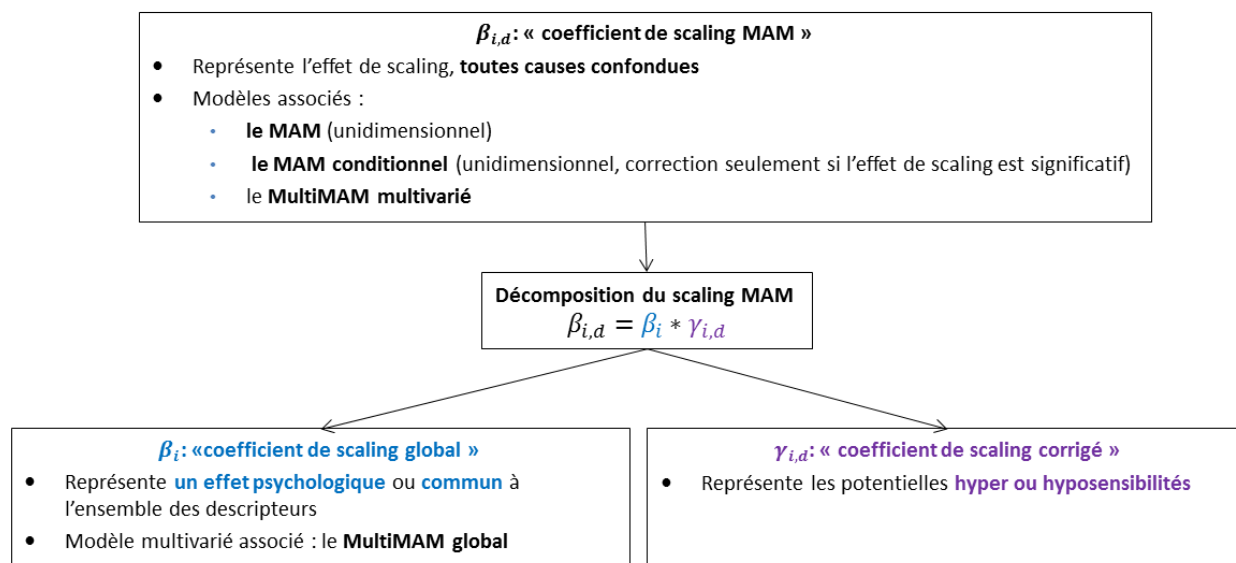


FIGURE 22 : SYNTHÈSE DES MODÈLES ET DES COEFFICIENTS DE SCALING

Les méthodes proposées ont été incluses dans des bibliothèques R disponibles sur www.timesens.com (MAMCAP, MultiMAM), permettant de réaliser des tables CAP ou MAM-CAP, et d'obtenir les résultats du modèle MultiMAM.

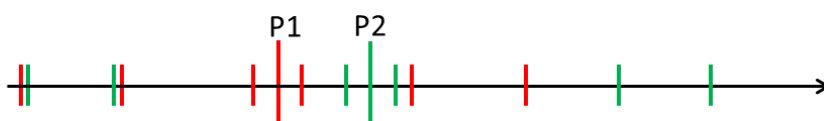
B. DES CARTES ADAPTEES AUX MODELES STATISTIQUES PROPOSES

I. LA CVA

1. CONCEPT

L'ACP est basée sur le tableau des moyennes des notes obtenues par les produits, elle ne prend donc pas en compte la variabilité des notes des sujets autour de cette moyenne. La Figure 23 illustre deux situations où les moyennes des produits ont la même variabilité, mais où la variabilité des notes des sujets n'est pas la même : on n'aboutit pas aux mêmes conclusions.

Situation 1: Forte variabilité des sujets impliquant que P1 et P2 ne sont pas différents



Situation 2: Faible variabilité des sujets impliquant que P1 et P2 sont différents



Deux produits ont été notés par 6 sujets dans deux situations.
Les lignes courtes représentent les notes des sujets, les grandes lignes représentent la moyenne

FIGURE 23 : ILLUSTRATION DE LA DIFFERENCE ENTRE VARIABILITE ET DISCRIMINATION

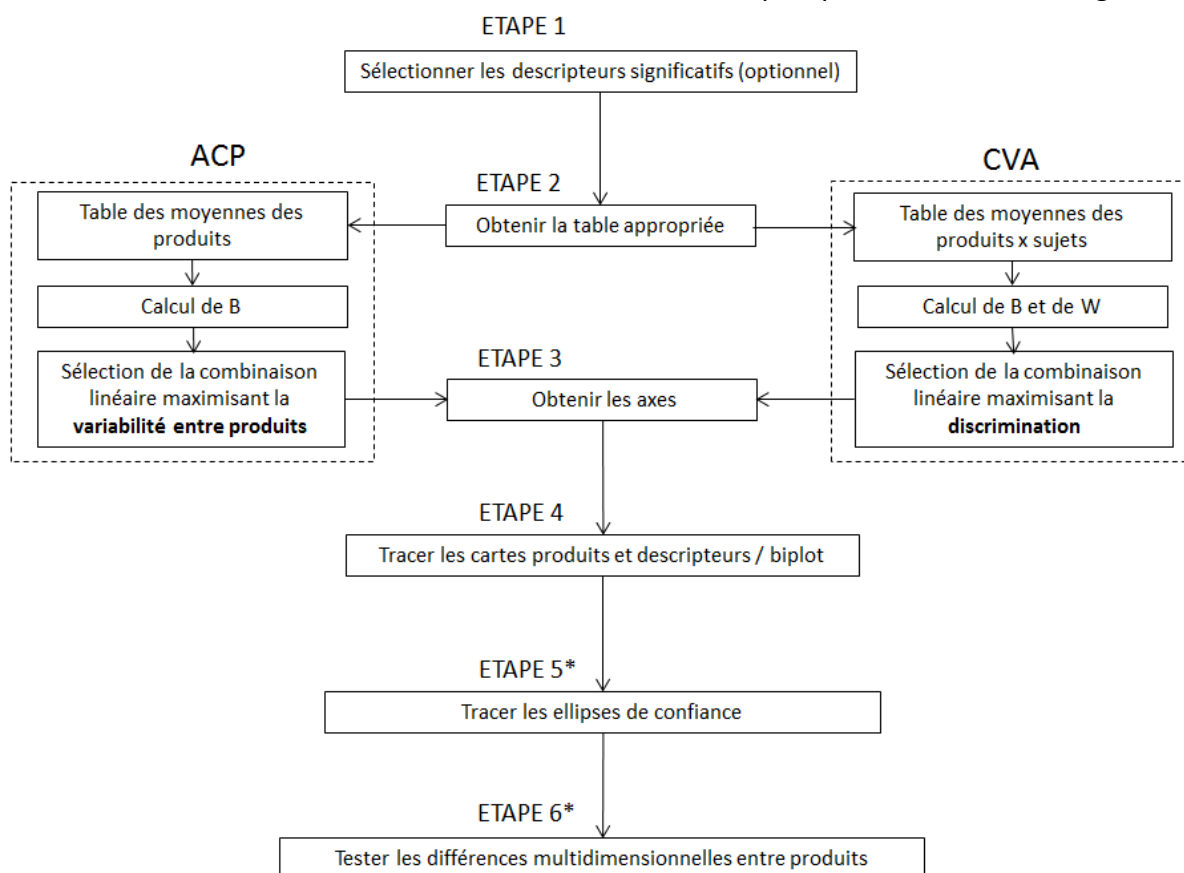
La CVA, proposée par Porcherot and Schlich (2000) pour l'analyse du profil sensoriel est une méthode de cartographie des produits permettant de prendre en compte cette variabilité (Peltier, Visalli, & Schlich, 2015a) à travers le concept de **discrimination multidimensionnelle**. Ainsi, ce n'est plus *la variabilité moyenne entre les produits* qui est représentée sur la carte –comme pour l'ACP–, mais la **discrimination entre les produits**, basée sur un modèle statistique.

Nous avons vu lors de la partie précédente que la discrimination unidimensionnelle des produits était évaluée à l'aide du rapport $F = \frac{CM_{Produit}}{CM_{Produit \times Sujet}}$. L'équivalent multidimensionnel d'un tel F est la matrice de discrimination de l'effet produit $W^{-1}B$ où B est la matrice de covariance de l'effet produit et W la matrice de covariance de l'interaction.

La **différence entre variabilité et discrimination est le cœur de la différence entre ACP et CVA**. En effet, la CVA est basée sur la diagonalisation de la matrice de discrimination $W^{-1}B$ alors que l'ACP est basée sur la diagonalisation de la matrice B (cf. Chapitre 1, C.I.).

2. CONSTRUCTION D'UNE CVA

Obtenir une carte de CVA s'effectue suivant différentes étapes présentées dans la Figure 24.



* Ces étapes ne sont pas classiques en ACP mais sont pourtant réalisables par projection des notes des sujets en individus supplémentaires

FIGURE 24 : DIFFERENTES ETAPES DE CONSTRUCTION D'UNE CARTE SENSORIELLE

La première étape (optionnelle) d'une CVA consiste à ne **sélectionner que les descripteurs significatifs** à l'aide d'ANOVAs par descripteur. Cette étape présente l'avantage d'éviter un sur-paramétrage qui nuirait à l'inversion de la matrice W . Cette approche est critiquable, dans la mesure où des descripteurs non discriminants (mais à la limite de la discrimination) unidimensionnellement pourraient néanmoins avoir une importance multidimensionnelle, mais elle permet tout de même de retenir l'information essentielle : les descripteurs significatifs. La table contenant les notes (moyennées sur l'ensemble des répétitions) des produits*sujets est ensuite calculée, puis B , la matrice de covariance de l'effet produit, et W , la matrice de covariance de l'interaction sont également calculées. La **matrice de discrimination** $W^{-1}B$ est ensuite diagonalisée. On obtient ainsi les **axes** de la carte (vecteurs propres issus de la diagonalisation) qui représentent la **combinaison linéaire des descripteurs maximisant la discrimination**. Le nombre d'axes à analyser dans une CVA peut être obtenu par un test du nombre de dimensions significatives de l'espace produit (test de rapport de vraisemblance, Mardia, Kent, and Bibby (1989)). Il n'existe pas de tel test en ACP.

Les moyennes obtenues par les produits sont ensuite projetées sur ces axes, et les corrélations entre les descripteurs et les axes sont calculées. J. Gower, Lubbe, and Le Roux (2011) proposa d'adapter le biplot (introduit par Gabriel (1971) dans le cadre de l'ACP à la CVA. Comme pour l'ACP, le biplot permet à l'utilisateur de « projeter » directement un produit sur les descripteurs, donnant ainsi une idée de la moyenne obtenue par ce produit *pondérée par l'accord entre les sujets*. Les détails techniques pour produire ce biplot sont disponible dans Peltier et al. (2015a).

Pour déterminer si deux produits sont différents pour un seul descripteur, on utilise généralement un t test. L'extension multidimensionnelle du t test est un test T2 de Hotelling. Ce test peut être effectué sur l'ensemble des dimensions, sur l'espace des dimensions significatives ou sur les deux premières dimensions seulement. Les différences entre les produits sont testées par des tests T2 de Hotelling sur l'espace des dimensions significatives. Lorsque deux produits sont statistiquement égaux multidimensionnellement, la p-value du test de Hotelling est supérieure au risque alpha fixé en amont, et les produits sont reliés par un segment sur la carte de CVA.

Des ellipses de confiance représentant un intervalle de confiance bi-dimensionnel des moyennes des produits sont ensuite tracées. Ce point sera développé dans III.

3. EXEMPLE SUR LE JEU DE DONNEES CHEESES

La Figure 25 présente la CVA–biplot associée au jeu de données « cheeses », et montre comment projeter les ellipses des produits D et F sur les attributs Salted et Sticky. Le Tableau 21 montre les résultats des tests de Hotelling associés au même jeu de données.

	A	B	C	D	E	F
A	1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
B		1	<0.0001	0.0011	<0.0001	<0.0001
C			1	<0.0001	0.194	<0.0001
D				1	<0.0001	0.0011
E					1	<0.0001
F						1

TABLEAU 21 : P-VALUES DES TESTS DE HOTELLING ASSOCIEES AU JEU DE DONNEES CHEESES SUR LE PLAN DES DEUX PREMIERS AXES

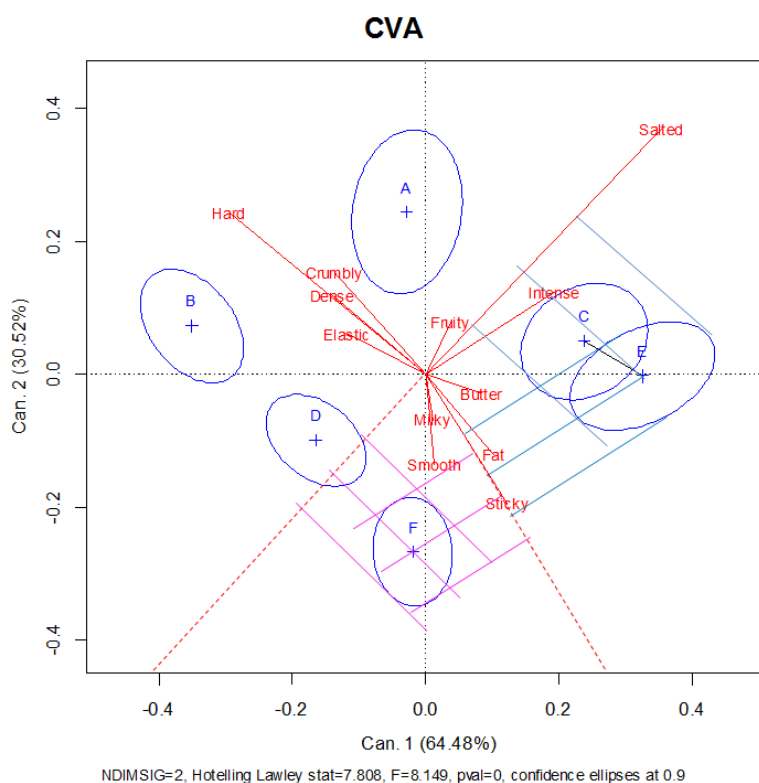


FIGURE 25 : CVA-BIPLLOT ASSOCIEE AU JEU DE DONNEES CHEESES

Les positions des produits et des descripteurs sont relativement semblables à celles observées dans l'ACP du chapitre 1. Les ellipses peuvent également être projetées sur les descripteurs (traits roses et bleus) pour donner des indications sur les intervalles de confiance associés à la moyenne de chaque produit par descripteur. Les produits C et E ne sont pas multidimensionnellement différents d'après le test de Hotelling (case en gras dans le Tableau 21, $p=0.194$). Les produits E et F semblent avoir obtenu des moyennes similaires sur le descripteur Sticky, et des moyennes très différentes sur le descripteur Salted.

4. DISCUSSION

Alors que les axes de l'ACP sont orthonormés, les axes de la CVA sont W -orthonormés (ce qui signifie que les axes ne sont pas systématiquement perpendiculaires). Cependant, il est d'usage de les tracer orthogonaux (comme en analyse discriminante) dans un souci de lisibilité du graphique.

La CVA présente un inconvénient majeur : elle utilise l'inversion de la matrice de désaccord W . La question de l'inversibilité de cette matrice se pose donc, ainsi que celle de son conditionnement (stabilité de son inverse). Lorsque la matrice est non inversible ou mal conditionnée, il est déconseillé d'effectuer une CVA.

Les fonctions CVA ont été incluses dans la librairie CVApack présentée dans Peltier et al. (2015b). Cette cartographie permet bien de répondre à la problématique présentée en introduction : elle est basée sur le modèle multidimensionnel permettant de calculer la

discrimination des produits (et prend donc en compte la variabilité des notes des sujets autour des moyennes).

II. LA MAM-CVA

1. PRINCIPE

La CVA a ensuite été adaptée au modèle MAM avec la MAM-CVA. Cette carte présente l'avantage de prendre compte les différents effets de scaling présentés précédemment (global ou multivarié, voir chapitre 3. A, II.). Elle est basée sur la diagonalisation de la matrice de discrimination appropriée (prenant en compte l'effet de scaling) comme indiqué Figure 26.

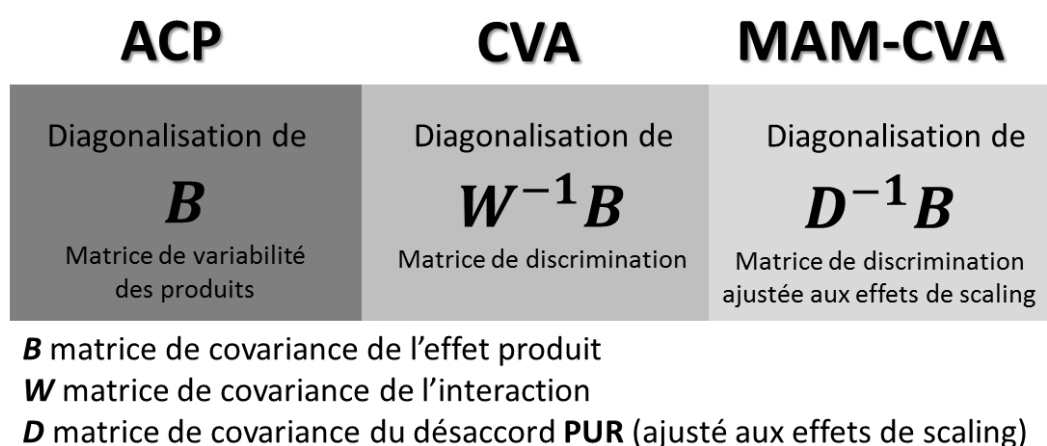


FIGURE 26 : PRINCIPES DES DIFFERENTES CARTOGRAPHIES DES PRODUITS

La MAM-CVA est construite exactement suivant les mêmes étapes que la CVA (présentées Figure 24), à cela près que la matrice de covariance de l'interaction W est remplacée par la matrice de covariance de désaccord pur D . Cette matrice peut prendre plusieurs valeurs en fonction du modèle choisi. Avec le modèle multiMAM-global, cette matrice représente le désaccord épuré de l'effet de scaling global (commun à tous les descripteurs). Avec le modèle MultiMAM-multivarié, cette matrice représente le désaccord épuré de toute source de scaling. Avec le modèle MultiMAM conditionnel, la matrice représente le désaccord épuré de toute sorte de scaling, seulement quand l'hétérogénéité du scaling est significative pour le panel.

La MAM-CVA est présentée en détails dans Peltier et al. (2015e).

2. DISCUSSION

Comme la CVA, la MAM-CVA est également soumise à des problèmes éventuels d'inversion de matrices et de conditionnement. On peut noter que les problèmes présentés dans les tests de performances multidimensionnelles (paradoxe de la MANOVA) n'ont pas

d'incidence sur la MAM-CVA produite. En effet, les problèmes détectés dans la MANOVA concernent uniquement le calcul de la p-value. Il reste possible de calculer la matrice de discrimination adéquate.

Par conséquent, nous conseillons l'utilisation de la MAM-CVA globale qui permet de prendre en compte le scaling sans le sur-corriger.

III. A PROPOS DES ELLIPSES DE CONFIANCE

1. INTRODUCTION

Dans ce travail, les ellipses de confiance représentent une extension multidimensionnelle de l'intervalle de confiance $[\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}; \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}]$ (avec \bar{x} moyenne observée et σ écart-type observé sur un échantillon de taille n).

Lorsqu'on utilise un modèle d'ANOVA pour tester les différences entre les moyennes de produits, il est fortement recommandé de remplacer dans la formule de l'intervalle de confiance l'écart type « global » σ , par la racine carrée du carré moyen du terme d'erreur du modèle. Par exemple, si le modèle d'ANOVA produit**sujet* avec effet *sujet* aléatoire est utilisé, $F_{Prod} = \frac{MS_{produit}}{MS_{interaction}}$ est calculé, et on utilise les carrés moyens de l'interaction $MS_{interaction}$ comme terme d'erreur. L'intervalle de confiance associé est alors le suivant :

$[\bar{x} - z_\alpha \frac{\sqrt{MS_{interaction}}}{\sqrt{n}}; \bar{x} + z_\alpha \frac{\sqrt{MS_{interaction}}}{\sqrt{n}}]$. Cet intervalle est théoriquement plus grand que l'intervalle de confiance calculé à partir de MS_{erreur} parce qu'en l'absence d'interaction significative, le carré moyen de l'interaction a une espérance mathématique égale à la variance théorique de l'erreur du modèle.

Les ellipses de confiance représentent l'extension bi-dimensionnelle de l'intervalle de confiance. Ainsi, les ellipses de confiance classiques (Saporta, 1986) sont représentées à l'aide d'un estimateur de la matrice de covariance des points autour du point moyen (extension multidimensionnelle de la notion d'écart-type). La démarche habituelle pour calculer l'ellipse de confiance d'un produit fixé est la suivante :

- projection des notes (moyennes sur les éventuelles répétitions) des sujets (brutes) sur la carte, moyennées par répétition,
- calcul de la matrice de covariance et des coordonnées des points obtenus,
- utilisation de la matrice de covariance et de la loi bi-normale pour déterminer l'ellipse de confiance associée (avec un seuil de confiance α).

La moyenne a alors une probabilité de $(1 - \alpha)$ d'être dans la zone de confiance construite. Il existe un autre type d'ellipse de confiance que nous appellerons ici « ellipse individuelle ». Ce type d'ellipse indique une zone dans laquelle les notes des sujets se situent avec une probabilité de $(1 - \alpha)$ (Figure 27).

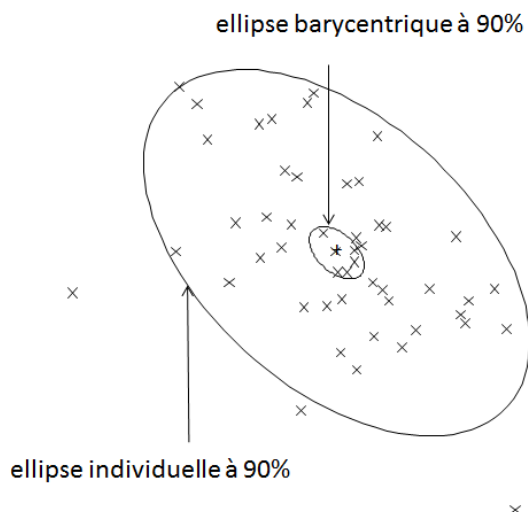


FIGURE 27 : EXEMPLE D'ELLIPSE DE CONFIANCE

2. UNE NOUVELLE METHODE POUR CONSTRUIRE DES ELLIPSES DE CONFIANCE

La CVA est basée sur un modèle multidimensionnel sous-jacent dans lequel l'effet produit est testé contre l'interaction pour obtenir la meilleure discrimination possible. Il paraît cohérent de tester également les différences entre les produits deux à deux contre ce même terme d'interaction, et de représenter la zone de confiance correspondant à cette interaction, à l'image des intervalles de confiance unidimensionnels.

Ainsi, pour construire des ellipses de confiance, nous proposons de projeter uniquement l'effet produit et le terme d'interaction sur la carte. Pour cela, chaque note attribuée par un sujet pour un produit et un descripteur (moyennée sur les répétitions) est décomposée en la somme de son effet moyen, son effet sujet, son effet produit et son interaction. Pour effectuer la correction, on considère ensuite les notes corrigées, constituée uniquement de la somme de l'effet moyen, l'effet produit et son interaction (et donc sans l'effet sujet). Ce sont ces notes corrigées qui sont projetées sur la carte et sur laquelle seront basées les ellipses de confiance. Les ellipses résultant de ces projections représenteront ainsi le désaccord des sujets autour de la moyenne de ce produit, épuré de l'effet sujet (c'est-à-dire que le calcul du désaccord prend en compte le fait que certains sujets notent plus haut que d'autres). La carte obtenue sera nommée CVA avec l'effet sujet non projeté, et mis à part la taille des ellipses, sera la même que la CVA.

Cette approche peut être étendue à l'utilisation du MultiMAM et à la MAM-CVA : dans ce cas il s'agit de projeter uniquement l'effet produit et le terme de désaccord pur, sans l'effet de scaling (Tableau 22).

Carte	Pour la construction des ellipses, projection de :
CVA	L'effet produit, de l'effet sujet et de l'interaction
CVA avec effet sujet non projeté	L'effet produit et de l'interaction
CVA multivariée	L'effet produit et du désaccord pur (hors scaling multivarié)
CVA globale	L'effet produit et l'effet du désaccord pur (hors scaling global)

TABLEAU 22 : DIFFERENTS TYPES D'ELLIPSES POUR DIFFERENTES CARTES

Le test de Hotelling a lui aussi été adapté afin de tester les différences inter-produits contre le désaccord pur. Pour cela, au lieu d'effectuer le test de Hotelling sur les données brutes, on effectue le test sur les données auxquelles on a préalablement soustrait l'effet de scaling et l'effet sujet (Peltier et al., 2015e).

3. EXEMPLE

Les cartes présentées Figure 28 sont obtenues à partir du jeu de données « cheeses ».

Comme prévu, les cartes de la CVA et la CVA avec l'effet sujet non projeté sont les mêmes, seule la taille des ellipses change (dans la CVA avec l'effet sujet non projeté, l'effet sujet a été retiré avant que les points ne soient projetés). Les MAM-CVA sont quant à elles différentes de la CVA : les plans sont différents, et les positions des produits et des descripteurs sont légèrement différentes. Le F de la MAM-CVA multivariée est plus grand que celui de la MAM-CVA globale, lui-même plus grand que celui de la CVA classique. On observe bien que les ellipses sont de plus en plus petites. Les nouvelles ellipses et le test de Hotelling permettent de discriminer les produits C et E dans les MAM-CVA et CVA avec l'effet sujet non projeté, ce qui n'était pas le cas dans la CVA de base.

4. DISCUSSION

Un avantage de la CVA est qu'elle permet de produire des ellipses de confiance en accord avec le modèle statistique utilisé pour tester les différences entre produits. Les ellipses que nous utilisons ici représentent la distribution du désaccord des sujets pendant notre étude. Elles répondent à la question « dans mon étude, quelle a été la variabilité (totale, ou du désaccord) de *mes* juges autour de ce produit ? ». Elle est ainsi en lien direct avec le modèle mis en place. Un autre type d'ellipse existe, basé sur des techniques de rééchantillonnage (bootstrap). Dans ce cas, ces ellipses répondent à la question « si j'avais eu un autre panel, aurais-je eu les mêmes résultats ? ». Il pourrait être envisageable de retravailler les ellipses de désaccord pur avec des techniques de bootstrap.

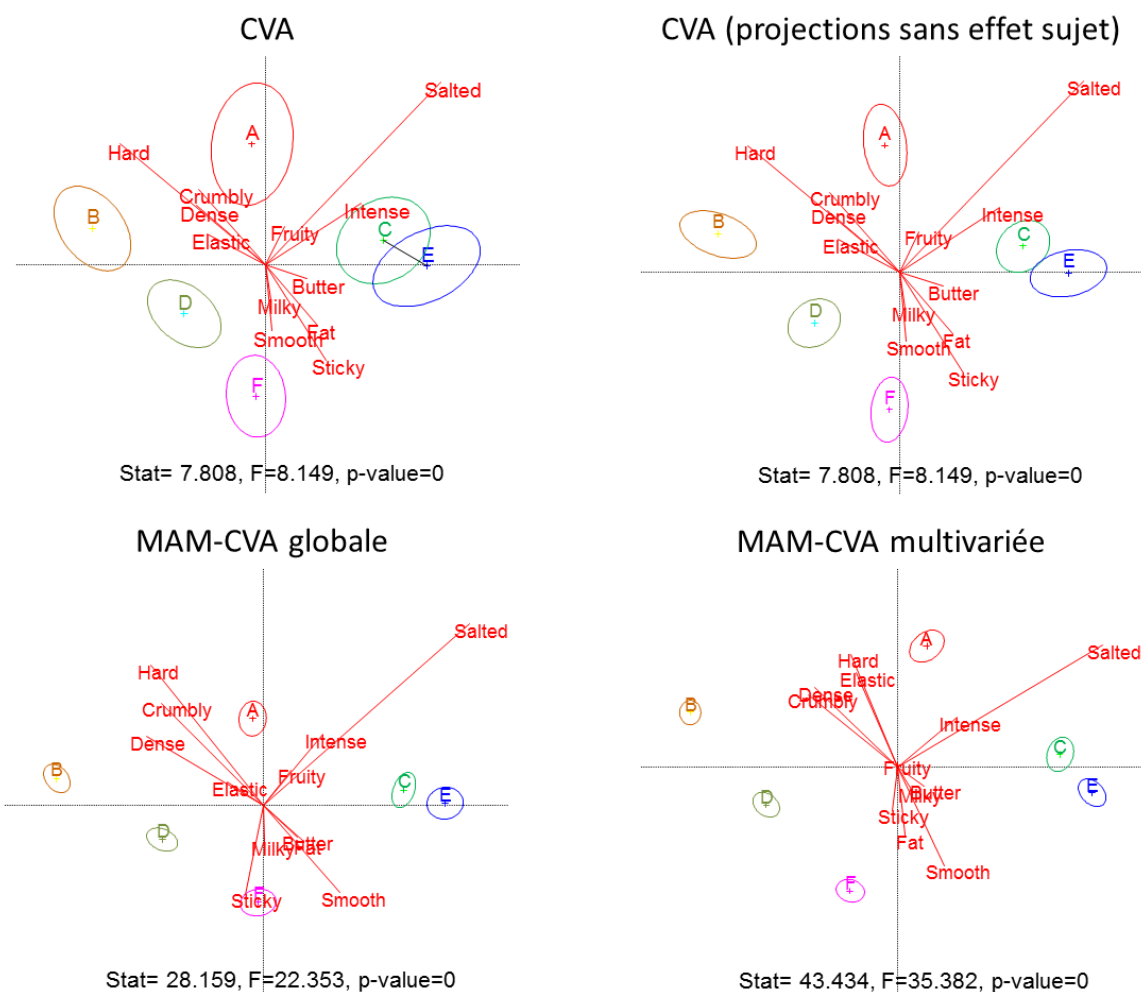


FIGURE 28 : DIFFERENTS TYPES DE CVA SUR LE JEU DE DONNEES CHEESES

IV. BILAN SUR LES APPORTS SUR LA CARTOGRAPHIE DES PRODUITS

La CVA répond précisément au problème posé dans l'introduction car elle permet d'effectuer une carte de la **discrimination des produits** plutôt qu'une carte de la variabilité des produits (comme le fait l'ACP). La CVA se base sur un **modèle statistique**, et a été adaptée au MultiMAM (MAM-CVA). La MAM-CVA permet ainsi de prendre en compte l'effet de scaling (global ou multivarié). Les ellipses de confiance ont été repensées afin d'être en accord avec le modèle multidimensionnel choisi.

Nous recommandons **l'utilisation de la carte MAM-CVA globale**. En effet, cette carte présente l'avantage de corriger l'effet de scaling, mais uniquement quand il est présent sur la plupart des descripteurs (et donc ne pas sur-corriger un effet de scaling).

Deux librairies R ont été implémentées afin de construire ces cartes : CVA et MultiMAM (disponibles sur www.timesens.com)

C. CONCLUSION DU CHAPITRE 3

Ce chapitre présente différentes méthodes théoriques pour améliorer l'analyse du profil sensoriel (analyse des performances et cartographie). Les méthodes présentées dans ce chapitre permettent de prendre en compte l'effet de scaling dans les cartographies ou les tests de performances. Des outils pratiques (bibliothèques R) ont également été fournis afin de permettre à tout animateur de panel de les utiliser (bibliothèques MAMCAP, MultiMAM, CVA). L'aide à l'utilisation de ces packages est présentée en Annexe 3.

Ces méthodes doivent à présent être utilisées sur un grand nombre de jeux de données réelles, afin d'être éprouvées et validées. De telles études pourront permettre de :

- quantifier les différences entre les nouvelles et les anciennes cartes,
- mettre en évidence les paradoxes de la MANOVA évoqués en A.II.2,
- obtenir des indications sur les performances des panels et des sujets, uni- et multidimensionnelles (distribution, lien avec les caractéristiques des sujets ou les modalités sensorielles notées...),
- observer l'effet de la décomposition du scaling en deux composantes psychologique et physiologique.

CHAPITRE 4 :

META-ANALYSES DE LA SENSOBASE

Dans ce chapitre, des méta-analyses de la SensoBase permettront de tester sur un grand nombre de jeux de données les nouvelles méthodes statistiques présentées dans le chapitre 3. Toutes nos méta-analyses seront effectuées à l'aide de la méthodologie mise en place dans le chapitre 2. Les rapports complets seront résumés dans ce chapitre et fournis en annexes pour plus de précisions. Certaines méta-analyses ont été intégrées dans le logiciel TimeSens afin que l'utilisateur puisse les reproduire avec son propre sous-ensemble de jeux de données.

Tout d'abord, **différentes méthodes de cartographie de profils sensoriels seront comparées** : l'ACP et la CVA, l'ACP et la MAM-CVA (globale et multivariée), l'ACP et l'ACP verticale (une autre alternative à l'ACP). Les **performances du panel et des panélistes** seront ensuite analysées. Elles seront quantifiées puis mises en regard avec les caractéristiques des sujets ou des descripteurs. Enfin, les questions de **l'utilité des répétitions** et du nombre de sujets supplémentaires équivalents à une répétition supplémentaire seront étudiées.

A. COMPARAISON DE CARTES

I. METHODOLOGIE

L'objectif de cette partie est de comparer des cartes d'ACP et de CVA (plan des produits et cercle des corrélations) de la manière la plus exhaustive possible. Pour cela, **les différentes caractéristiques d'une carte** ont été listées : la **configuration des produits**, l'**interprétation sensorielle** déterminée par la position relative des descripteurs, le **pouvoir de discrimination** multidimensionnelle et les **comparaisons statistiques des produits 2 à 2** (obtenues par l'intersection des ellipses de confiance). Afin de comparer l'ensemble de ces éléments, différents critères ont été établis.

1. UN CRITERE DE COMPARAISON DES CONFIGURATIONS DES PRODUITS

Les configurations des produits peuvent être comparées à l'aide d'un coefficient RV (ce critère sera nommé par la suite **rvCoef**). Le coefficient RV (Escoufier, 1973) est un coefficient de corrélation multidimensionnel, invariant par rotation, dilatation ou symétrie axiale des points. Un coefficient RV proche de 1 implique que les configurations des produits sont similaires. Inversement, un coefficient proche de zéro indique que les deux configurations ne sont pas similaires.

2. DEUX CRITERES DE COMPARAISON DES INTERPRETATIONS SENSORIELLES

Sur une carte d'ACP ou de CVA, les positions des produits sont interprétées en fonction des directions des descripteurs. Afin de comparer les interprétations de deux cartes, les deux configurations des produits sont superposées (par rotation procrustéenne, (J. Gower, 1975)). Puis la même transformation est appliquée aux descripteurs, afin de pouvoir

comparer leurs directions. Pour chaque descripteur, l'angle (en degrés) entre la flèche (représentant ce descripteur) d'une carte et celle de l'autre carte sont mesurés (Figure 29).

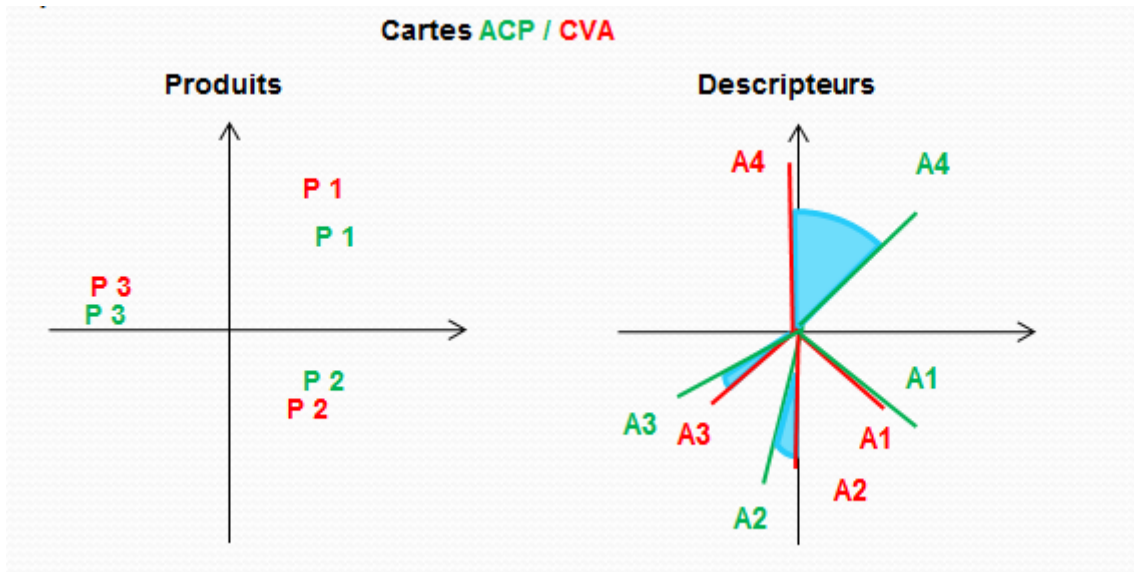


FIGURE 29 : ANGLES ENTRE LES DESCRIPTEURS DE L'ACP ET LA CVA (APRES ROTATION PROCRUSTEENNE)

Deux critères sont ensuite calculés pour chaque jeu de données :

- L'angle maximal entre les descripteurs issus de la CVA et les descripteurs issus de l'ACP (**maxAngle**) indiquant la différence d'interprétation maximale.
- L'angle moyen (**avgAngle**) entre ces descripteurs ACP et CVA indiquant la différence d'interprétation moyenne.

3. UN CRITERE DE COMPARAISON DES INTERSECTIONS DES ELLIPSES DE CONFIANCE

Pour chaque carte et chaque paire de produits, le test T2 de Hotelling (Chapitre 3, B, I. 2.) peut être effectué pour déterminer si les produits sont multidimensionnellement différents ou non. Ainsi, une p-value significative indique que les produits sont significativement différents tous descripteurs confondus. En général, les résultats de ce test coïncident avec l'interprétation des ellipses de confiance : si deux ellipses sont disjointes (ou non) alors le test de Hotelling sera généralement significatif (ou non). Le critère retenu au niveau du jeu de données a été la **proportion de paires de produits dont le diagnostic (égaux ou différent) était le même pour les deux cartes (PctId)**. Eventuellement, la proportion de paires de produits significativement différentes dans un cas et pas dans l'autre pourra également être calculée – par exemple la proportion de paires de produits détectées différentes pour la CVA et pas pour l'ACP et inversement.

4. UN CRITERE DE COMPARAISON DU POUVOIR DISCRIMINANT MULTIDIMENSIONNEL

Le rapport des F de la MANOVA (**F-Ratio**) a été calculé pour quantifier l'amélioration de la discrimination entre l'ACP et la CVA (F_{CVA}/F_{ACP}). Théoriquement, ce rapport devrait être systématiquement plus grand que 1. Ce F-ratio peut prendre des valeurs très élevées. Par conséquent, lorsqu'on aura plusieurs valeurs de F-ratio, on préférera analyser la médiane de la distribution plutôt que la moyenne (très sensible aux valeurs extrêmes).

II. ACP CONTRE CVA

L'ACP et la CVA ont été comparées sur 422 jeux de données de profil sensoriel équilibrés (Annexe 1.B) (Peltier et al., 2015b).

Cette étude a montré que **l'ACP et la CVA étaient relativement semblables** sur la plupart des critères. Le RV moyen entre les configurations des produits étaient de 0.91, l'angle maximal (respectivement moyen) entre deux descripteurs était de 33 degrés en moyenne (respectivement 6 degrés en moyenne). La médiane du **F-ratio** est de 1.8. Les différences entre produits sont les mêmes dans 91% des cas, et 4% (respectivement 5%) des paires de produits ont été discriminées dans la CVA et pas dans l'ACP (respectivement dans l'ACP et pas dans la CVA).

Il est à noter qu'en faisant les analyses par modalité sensorielle, les cartes sont encore plus semblables. Quelques jeux de données présentent cependant des différences entre leurs représentations ACP et CVA. **Ces différences peuvent être expliquées par un grand nombre de produits et une forte complexité de l'espace produit** (grand nombre de descripteurs significatifs). Pour ces jeux de données, **construire les cartes par modalité sensorielle des descripteurs** permet de limiter les différences entre ACP et CVA.

Notons que la comparaison de cette étude a été effectuée sur le plan 1-2, mais aussi sur l'espace au nombre de dimensions significatives de la CVA.

La fréquence des problèmes de conditionnement des matrices utilisées dans la CVA a également été étudiée. Environ 80% des jeux de données n'ont pas de problème de conditionnement, et 3% ont un réel problème de conditionnement (pouvant conduire à des cartes qui ne seraient pas stables). Par conséquent les **problèmes de conditionnement de matrices sont mineurs** dans la construction des CVA pour les jeux de données de profil sensoriel.

Alors que la comparaison précédente concerne les cartes constituées de deux graphiques (une carte produit et le cercle de corrélations associé), la comparaison des biplots de l'ACP et de la CVA (graphique unique où figurent descripteurs et produits, qui diffère de la simple superposition des deux graphiques précédents) a également été effectuée (Rapport R1, Annexe 2.B), on obtient des résultats similaires, même si les différences moyennes entre les

angles des descripteurs sont légèrement plus grandes (44 degrés pour maxAngle et 16 degrés pour avgAngle).

III. ACP CONTRE CVA-MAM

La comparaison de l'ACP à la CVA sans l'effet sujet (effet sujet retiré des ellipses), à la MAM-CVA multivariée (prenant en compte l'effet de scaling MAM) et à la MAM-CVA globale (prenant uniquement en compte l'effet de scaling global, commun à tous les descripteurs) a été effectuée sur les mêmes critères que précédemment (Rapport R2, Annexe 2.B). Les résultats sont présentés dans le Tableau 23.

Critères de comparaison	ACP-CVA (article)	ACP- CVA sans effet sujet	ACP MAM-CVA globale	ACP-MAM-CVA multivariée
Nombre de jeux de données	422	379	379	377
RV coef	0.91	0.92	0.95	0.92
Max Angle	33	26	23	25
Avg Angle	6	7	6	7
Pourcentage de paires de produits discriminées en ACP mais pas en CVA	5%	0.6%	0.3%	0.3%
Pourcentage de paires de produits discriminées en CVA mais pas en ACP	4%	17%	18%	19%
PctId : pourcentage de paires de produits au même diagnostic 'égaux/différents' en ACP et en CVA	91%	82%	81%	80
F Ratio (med)	1.8	1.79	1.3	1.13

TABLEAU 23 : VALEURS DES CRITERES DE COMPARAISON ENTRE ACP ET CVA, CVA SANS PROJECTION DE L'EFFET SUJET ET MAM CVA

On observe qu'il y a de **fortes similitudes** en ce qui concerne les **configurations produits** (rvCoef) et les **interprétations sensorielles** (maxAngle et avgAngle) de la MAM-CVA et l'ACP. En revanche, les tests de Hotelling et les ellipses de confiance donnent des résultats différents en fonction de la carte choisie. Lorsqu'on retire l'effet sujet des points projetés avant de construire les ellipses de confiance, le **pourcentage de paires discriminées dans la CVA et non discriminées dans l'ACP s'élève à 17% contre 4% seulement dans le cas des ellipses classiques**. Ce chiffre augmente encore avec la MAM-CVA (18% pour la MAM-CVA globale et 19% pour le multivarié).

IV. APPLICATION DE LA METHODOLOGIE DE COMPARAISON DE CARTES A L'ACP VERTICALE

L'ACP verticale proposée par Dettmar, Plattner, and Vetter (2014) est basée sur le constat suivant : l'ACP est effectuée sur le tableau des moyennes de produits, contenant souvent plus de variables (descripteurs) que d'individus (produits), ce qui n'est pas recommandé. Pour pallier ce problème, Dettmar et al. (2014) propose de remplacer le tableau produit par le tableau de données produit x sujet x répétition, centré par sujet et divisé par l'écart type du panéliste. Cet écart type a été obtenu en calculant la variance du panéliste pour chaque descripteur, puis la moyenne des variances, et finalement la racine de cette quantité.

Nous appliquons ici la méthodologie de comparaison de cartes pour l'ACP à l'ACP Verticale. Deux critères supplémentaires ont été ajoutés à la méthodologie de comparaison initiale (ACP/CVA) pour donner des indications sur la dimensionnalité de l'espace étudié.

- Beta : $(= 1 + 2 \frac{\sum_{i < j}^N \lambda_i \lambda_j}{\sum_i^N \lambda_i^2})$ où λ_i sont les valeurs propres de l'ACP. Ce coefficient a été présenté par Schlich (1996)
- N : représente le nombre d'axes requis pour représenter 80% de la variance ; ce pourcentage a été défini arbitrairement, il est cependant couramment utilisé en ACP pour choisir le nombre d'axes à représenter.

Comme le nombre de dimensions choisies dépend du nombre de produits, l'étude de dimensionnalité a été ensuite restreinte aux jeux de données pour lesquels le nombre de produits était plus grand que le nombre de descripteurs significatifs.

La comparaison a été effectuée sur 441 jeux de données (Rapport R3, Annexe 2.C).

Les résultats ont montré que l'ACP et l'ACP verticale sont généralement relativement similaires, mais on a observé des changements dans les **directions des descripteurs** (17 degrés entre les descripteurs de l'ACP et ceux de l'ACP verticale en moyenne). L'ACP donne une carte avec une **discrimination multidimensionnelle légèrement meilleure** (amélioration du F de 6% en moyenne, et obtention de 2% de paires de produits discriminés supplémentaires). L'ACP verticale **utilise deux fois plus dimensions que l'ACP en moyenne**. Une dimensionnalité élevée indique un espace produit riche et complexe, et l'ACP Verticale permet de mieux représenter cette richesse, même si c'est au détriment d'une différenciation moindre des produits.

V. DISCUSSION ET BILAN SUR LES COMPARAISONS DE CARTES

La méthodologie mise en place a montré son efficacité pour différentes comparaisons. Elle pourrait être améliorée dans le cas des biplots en prenant en compte la taille des flèches correspondant aux descripteurs sur la carte, par exemple à l'aide d'un rapport (ou une différence) des tailles des flèches entre deux descripteurs.

Cette méthodologie a permis de comparer l'ACP et la CVA (à deux facteurs, CVA sans prendre en compte l'effet sujet dans les ellipses, CVA-MAM global et CVA-MAM-multivariée) et de mettre en évidence l'intérêt pratique d'utiliser notre méthode de construction des ellipses de confiance (ellipses disjointes pour 15% supplémentaires de paires de produits).

Il existe plus de différences entre l'ACP et l'ACP verticale qu'entre l'ACP et la CVA. L'ACP verticale permet d'utiliser plus de dimensionnalité, mais elle montre moins de différences entre les produits. En termes sensoriels, cela implique que l'espace produit défini est plus complexe dans l'ACP Verticale.

Globalement, l'ensemble des comparaisons a montré que les cartes (interprétation sensorielle et configuration produits) étaient le plus souvent relativement semblables, et cela quelle que soit la méthode choisie. Cela signifie que l'apport de nouvelles subtilités dans le modèle (avec le MAM) ou la représentation choisie (deux cartes ou biplot) ne change pas fondamentalement les conclusions de l'étude. L'animateur de panel effectuant des ACP (théoriquement moins adaptées que les autres cartes proposées ici) n'aurait donc en général pas interprété ses données différemment avec une CVA (sauf sur les différences entre produits et la taille des ellipses de confiance). Cependant, avec l'ACP des moyennes, les animateurs de panel ne comparent jamais les produits statistiquement au niveau multidimensionnel et restent sur une approche visuelle. Les CVA (quel que soit le modèle choisi) sont plus adaptées aux données de profil sensoriel, et sont cohérentes avec le modèle d'ANOVA choisi pour détecter des différences entre les produits. Ainsi, nous conseillons d'utiliser la carte MAM-CVA, qui prend en compte le maximum d'effet (scaling, sujet) sans sur-corriger les notes.

B. PERFORMANCES

Cette partie comprend la méta-analyse des performances unidimensionnelles du panel et des panélistes (distribution des indicateurs, effet des caractéristiques des sujets sur leurs performances) et l'étude des performances multidimensionnelles (utilisation de la MANOVA, scaling global).

I. PERFORMANCES UNIDIMENSIONNELLES

1. PERFORMANCES DU PANEL

Lors de la période d'entraînement du panel, l'étude des performances de ce panel est centrale : c'est elle qui permet de décider si le panel est prêt à passer à la phase de mesure. L'étude des performances des panels de la SensoBase (Rapport R4, Annexe 2.D) vise à :

- Etudier la distribution des indicateurs de performances du panel.
- Comparer les performances du panel en fonction du modèle choisi (à deux facteurs ou MAM)
- Chercher un lien éventuel entre les différentes performances et la modalité sensorielle du descripteur.
- Chercher des corrélations éventuelles entre les différentes performances.

Les performances unidimensionnelles du panel ont été étudiées grâce à une méta-analyse menée sur 177 jeux de données (équilibrés, à échelle continue). Pour cela, les tests de la table MAM-CAP ont été effectués, et les pourcentages de descripteurs significatifs pour les différents effets ont été calculés.

On a obtenu en moyenne une valeur de 1.25 pour les \sqrt{CME} (sur une échelle de 0 à 10). L'utilisation du MAM par rapport au modèle à deux facteurs avec effet sujet aléatoire a permis d'obtenir 4 % supplémentaire de significativité de l'effet produit et 20% d'accord du panel. Un effet de scaling a été observé pour 45% des descripteurs.

Des corrélations significatives entre les différentes performances ont été observées :

- Plus le panel met des notes élevées, plus le panel semble répétable.
- Plus les produits sont discriminés, plus l'hétérogénéité de scaling est forte.
- Plus la répétabilité du panel est faible, plus le désaccord entre les panélistes est fort.

Concernant l'impact de la modalité sensorielle sur les performances du panel, on a pu remarquer que les **descripteurs visuels et de texture ont un fort taux de discrimination** (plus de 80% de ces descripteurs ont bien un effet produit significatif) et sont notés haut sur l'échelle (à plus de 3.5 en moyenne). Les **descripteurs d'odeur sont les moins discriminants** (moins de 70% de ces descripteurs ont un effet produit significatif). La répétabilité est

maximale pour les descripteurs de saveur (\sqrt{CME} moyen de 0.86 contre des CME supérieurs à un pour les autres modalités sensorielles).

Mammasse (2012) avait trouvé des \sqrt{CME} moyens de 1.02 sur ses jeux de données. Nos \sqrt{CME} sont légèrement plus élevés, mais ils restent néanmoins dans la distribution obtenue.

Selon une étude réalisée sur 369 jeux de données, Brockhoff et al. (2015) avaient obtenu 52.4% des descripteurs avec interaction significative, 45.3% avec hétérogénéité du scaling. 32% ont montré un désaccord significatif (soit 20% de désaccord en moins avec le MAM). Nos résultats confirment cette première étude.

2. PERFORMANCES UNIDIMENSIONNELLES DU SUJET

L'étude de la SensoBase peut également permettre de répondre à la question du lien entre performances des panélistes et caractéristiques des panélistes (âge, genre, fait de fumer ou non) (Rapport R5, Annexe 2.E). Deux approches différentes ont été utilisées pour valider les résultats obtenus.

Une première approche a été effectuée en considérant **une observation par sujet contenu dans un jeu de données** (moyennée sur les descripteurs) pondérée par la représentativité des différentes caractéristiques dans le jeu de données, (pondération utilisée par Pineau (2006), cf. Annexe 2.E.). Pour expliquer les performances, elle considère le modèle à deux facteurs (jeu de données et caractéristique) avec effet jeu de données et interaction considérés comme aléatoires. Avec cette approche, le fait que les jeux de données soient déséquilibrés (pas le même nombre d'hommes et de femmes, ou de chaque catégorie d'âge dans chaque jeu de données) peut poser des difficultés dans les tests effectués (ANOVA ou les tests LSD).

Par conséquent, une seconde approche a été proposée en **considérant une observation par modalité de la caractéristique dans chaque jeu de données** (obtenue en moyennant les données par modalité au sein d'un jeu de donnée). Ces observations ont également été pondérées par la représentativité des différentes caractéristiques dans le jeu de données. Le modèle résultant est un modèle additif à deux facteurs (jeu de données et caractéristique).

La première approche permet de travailler sur un plus grand nombre de données, mais est basée sur un dispositif déséquilibré, alors que la seconde approche permet de travailler sur des dispositifs équilibrés mais avec moins d'observations. Finalement, pour chaque performance et chaque caractéristique, les deux analyses ont été effectuées et mises en commun pour déterminer d'éventuelles tendances. Cette étude a été effectuée sur 177 jeux de données équilibrés avec des échelles continues.

Les résultats obtenus montrent que le **genre n'impacte pas significativement les performances** (discrimination, accord ou répétabilité). Les femmes semblent avoir une (très)

légère tendance à disperser plus leurs notes que les hommes et à noter moins haut. Globalement, **les fumeurs ont tendance à moins discriminer, noter moins haut et disperser moins leurs notes** que les non-fumeurs. Les résultats les plus significatifs de l'étude ont été observés pour l'âge des panélistes. Les **seniors dispersent moins leurs notes et sont plus répétables** que le reste du panel. De plus, les panélistes plus âgés ont tendance à moins discriminer et les jeunes ont tendance à noter moins haut que le reste du panel.

A notre connaissance, il n'existait pas encore de résultats pratiques sur les fumeurs, mais les résultats obtenus valident des hypothèses théoriques présentes dans la littérature (Chapitre 1). Dans l'analyse de Pineau (2006), les catégories d'âge suivantes avaient été sélectionnées : -30, 30-45, +45. Il avait observé que la tranche d'âge 30-45 était celle qui discriminait le mieux. D'autre part, des tendances indiquaient un niveau d'accord plus faible pour les personnes de plus de 45 ans et un niveau de répétabilité moindre pour les personnes plus jeunes. Ces résultats sont plutôt en accord avec ce que nous avons trouvé dans notre méta-analyse, même si nos classes d'âge et notre base de données initiale ne sont pas les mêmes.

II. PERFORMANCES MULTIDIMENSIONNELLES

1. PARADOXE DE LA MANOVA

La question de l'extension multidimensionnelle des performances a posé des problèmes théoriques pouvant mener à des paradoxes. L'objectif de cette partie est de mettre en évidence l'existence pratique des paradoxes de la MANOVA présentés dans le chapitre 3, A.II. 2. Contrairement aux autres méta-analyses, celle-ci ne fait pas l'objet d'un rapport en annexe, les résultats seront présentés directement dans le texte.

La p-value associée au F de la MANOVA a été calculée pour le modèle à 2 facteurs, pour le Multi-MAM multivarié, pour le Multi-MAM global et pour le Multi-MAM conditionnel. Ainsi, nous avons sélectionné 124 jeux de données équilibrés avec des échelles continues (pour lesquels les tests liés aux 4 modèles fonctionnaient – pas de problème d'inversion des différentes matrices de désaccord). Ces jeux de données ont été extraits des 177 jeux utilisés pour les performances individuelles.

Les résultats obtenus sont les suivants : le modèle multivarié corrigeant systématiquement le scaling a un effet produit significatif ($p=0.05$) pour 87% des jeux de données, alors que pour le modèle à 2 facteurs qui ne corrige pas le scaling, l'effet produit est significatif pour 91% des jeux de données. Autrement dit, retirer l'effet de scaling fait diminuer la significativité ! On observe bien le paradoxe décrit dans le chapitre 3 (A.II.2). Les p-values obtenues avec le modèle MultiMAM-global sont encore moins significatives avec seulement 78% des jeux de données avec un effet produit significatif. Le modèle conditionnel permet bien, en revanche, d'améliorer la discrimination : 94% des jeux de données ont un F produit significatif.

La méta-analyse a donc montré que l'extension des tests de performances avec la MANOVA ne permet pas d'améliorer la discrimination multidimensionnelle des produits. Elle a montré l'importance de la perte de degrés de liberté dans la MANOVA pour la significativité de l'effet produit. Les tests de performances multidimensionnels n'ont donc pas été validés. Un compromis a été trouvé en ne corrigeant pas systématiquement l'effet de scaling, ni les degrés de liberté associés au terme de désaccord pur. Il a été montré que cette solution fonctionne pratiquement. Cependant, cette solution dépendante des descripteurs n'est pas applicable dans le cas de scaling global (où l'on travaille tous descripteurs confondus).

2. SCALING

Même si des tests multidimensionnels n'ont pas été mis en place, une nouvelle performance de scaling multidimensionnelle a été étudiée à l'aide du scaling global et du scaling corrigé. L'objectif de cette étude est de déterminer les caractéristiques de ces nouveaux coefficients de scaling (lien avec les caractéristiques des panélistes, impact de la modalité sensorielle...).

La décomposition du scaling en deux composantes a été effectuée sur 187 jeux de données (Annexe 1.D) (Peltier et al., 2015d). Une fonction a été proposée afin de « symétriser » le coefficient de scaling. En effet, on sera amené par la suite à calculer des moyennes de ces coefficients et à les analyser par ANOVA. L'idée intuitive sous-jacente à cette fonction est *qu'un panéliste qui disperse trois fois moins ses notes que le panel devrait être aussi éloigné de la moyenne (c'est-à-dire 1, la valeur n'indiquant pas d'effet de scaling) qu'un panéliste qui disperse trois fois plus ses notes que le panel*. Ainsi, nous avons défini une fonction f (Figure 30) dont l'expression est la suivante :

$$f: x \in [1; +\infty[\rightarrow x - 1$$

$$f: x \in]0; 1[\rightarrow 1 - \frac{1}{x}$$

$$f: x \in]-\infty; 0] \rightarrow NA$$

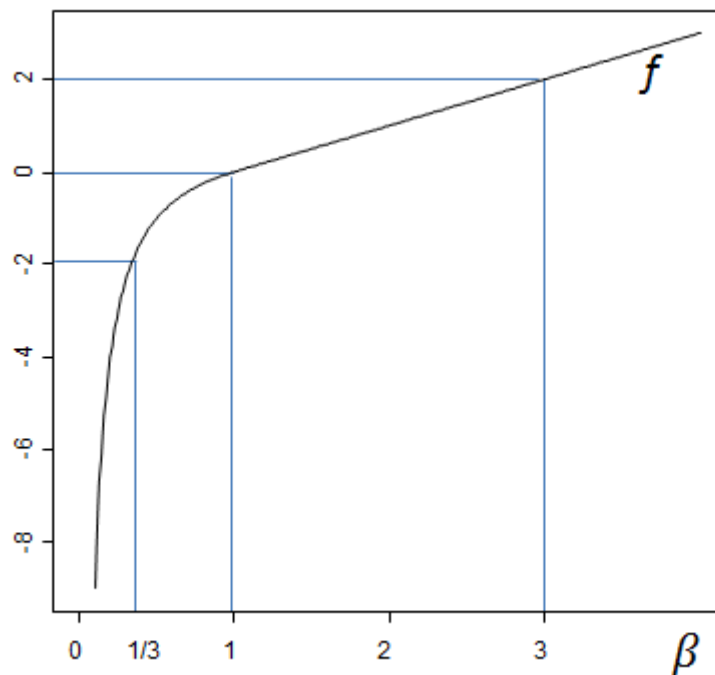


FIGURE 30: REPRESENTATION GRAPHIQUE DE LA FONCTION F

Cette fonction est continue, dérivable et inversible. Elle sera utilisée chaque fois qu'une moyenne de scaling sera calculée. Les méthodes utilisées pour répondre aux objectifs sont présentés dans Peltier et al. (2015d).

La méta-analyse a montré que les **personnes de plus de soixante ans et les fumeurs ont tendance à moins disperser leurs notes que les autres panélistes**. Ce résultat est cohérent avec les résultats de la bibliographie (Chapitre 1, 2) et les résultats obtenus en unidimensionnel. De plus, l'utilisation du scaling global a permis **de résoudre le problème de négativité des coefficients de scaling MAM** (16% de négativité pour le coefficient de scaling classique contre 0.9% pour le coefficient de scaling global). Cependant, en observant le coefficient de scaling global de sujets ayant participé à différentes études, nous n'avons pas noté de stabilité de ce coefficient de scaling au cours du temps. En d'autres termes, le coefficient de scaling global a semblé varier en fonction des jeux de données. Cela pourrait être expliqué par les différentes conditions expérimentales (consignes différentes, état d'esprit du sujet différent, changement de panel...).

Par ailleurs, la variance des scaling corrigés par descripteur détermine l'hétérogénéité des sensibilités pour ce descripteur. La méta-analyse a montré que cette hétérogénéité variait en fonction de la **modalité sensorielle du descripteur**. Ainsi, on a **observé plus de différences de sensibilité des sujets entre les descripteurs d'arôme, de goût et d'odeur** qu'entre les descripteurs visuels et de texture.

C. QUESTIONS GENERALES D'ANALYSE SENSORIELLE

I. INTERET DES REPETITIONS EN ANALYSE SENSORIELLE

La question du nombre de répétitions avait été posée par Mammasse (2012). Durant ce travail, des critères avaient été mis en place pour déterminer l'importance des répétitions. Cependant, ces critères ne prenaient pas en compte les cartes résultant du jeu de données avec une ou deux répétitions. De plus, depuis ces travaux, la SensoBase a été abondée par de nouveaux jeux de données. Nous avons cherché à quantifier les différences entre les jeux de données restreint à une répétition et ces mêmes jeux de données avec deux répétitions (Peltier et al., 2015c).

Pour arriver à cet objectif, nous avons comparé l'accord (unidimensionnel et multidimensionnel), la discrimination (unidimensionnelle et multidimensionnelle) et les CVA 1-2 de ces deux jeux de données.

Une méta-analyse sur 377 jeux de données (Annexe 1.F) a montré qu'en considérant uniquement le jeu de données restreint à la première répétition :

- la discrimination unidimensionnelle est la même pour 6 descripteurs sur 7,
- 87% des descripteurs significatifs ont un vecteur de moyennes des produits similaire à celui du jeu de données à deux répétitions (coefficient de corrélation >0.9),
- la discrimination multidimensionnelle est relativement équivalente (médiane du F-Ratio = 1.2),
- 90% des jeux de données ont une configuration produit similaire (RV >0.9),
- les cartes CVA 1-2 induites sont similaires (RV Coef=0.9, MaxAngle=36, AvgAngle=11, PctId=89%).

De plus, un petit nombre de produits ou un grand nombre de sujets dans l'étude semble améliorer les similitudes entre le jeu de données sans répétition et le jeu de données avec deux répétitions. En effet, on peut supposer que l'augmentation de la taille du panel ou la réduction de l'espace produit permet d'obtenir un jeu de données plus stable (et donc moins dépendant du nombre de répétitions).

Pour conclure, une fois le panel entraîné, les **répétitions ne semblent pas nécessaires pour obtenir des résultats satisfaisants en profil sensoriel**. Cependant, elles restent indispensables pour établir (ou vérifier) la répétabilité du panel et des sujets, et apportent quelques précisions supplémentaires. L'animateur de panel doit donc faire un compromis entre le coût de son expérimentation (augmentant avec l'ajout d'une répétition) et la qualité de son jeu de données (augmentant également avec l'ajout d'une répétition).

II. A COMBIEN DE SUJETS SUPPLEMENTAIRES EQUIVAUT UNE REPETITION SUPPLEMENTAIRE ?

La qualité d'un jeu de données peut être améliorée en ajoutant des répétitions, mais également en ajoutant des sujets au panel. Dans cette étude, nous avons cherché à mettre en relation le nombre de répétitions et le nombre de sujets dans l'étude. En d'autres termes, pour un jeu de données sans répétitions, à combien de sujets supplémentaires équivaut une répétition supplémentaire ?

Pour répondre à cette question, nous avons mis en place une étude présentée dans le rapport R6 (Annexe 2.F).

Cette étude a été effectuée sur 65 sous jeux de données (établis par ré-échantillonnages sur 13 jeux de données avec plus de 20 sujets) et a montré que le nombre de sujets supplémentaires équivalant à l'ajout d'une deuxième répétition valait entre 7 et 11 sujets (en moyenne). Cependant le nombre de sujets à ajouter reste extrêmement variable en fonction du jeu de données (écart type de 5). L'ajout d'une troisième répétition est quant à elle équivalente à l'ajout de 4 à 5 sujets.

D. CONCLUSION DU CHAPITRE 4

Dans ce chapitre, dix méta-analyses ont été effectuées (Tableau 24) selon la méthodologie mise en place dans le chapitre 2. Les rapports complets sont disponibles en Annexe 2.

Titre de la méta-analyse	Nombre d'études	Annexes
ACP contre CVA (two maps)	422	Peltier et al. (2015b), 1.B
ACP contre CVA (biplot)	387	Rapport R1, Annexe 2.A
ACP contre CVA-MAM (biplot)	387	Rapport R2, Annexe 2.B
ACP contre ACP verticale	441	Rapport R3, Annexe 2.C
Performances du panel	187	Rapport R4, Annexe 2.D
MANOVA et performances multidimensionnelles	114	Dans le texte
Performances des panélistes	177	Rapport R5, Annexe 2.E
Décomposition du scaling	187	Peltier et al. (2015d), Annexe 1.D
Répétition	377	Peltier et al. (2015c), Annexe 1.F
Répétition et nombre de sujets	222	Rapport R6, Annexe 2.F

TABLEAU 24 : SYNTHÈSE DES MÉTA-ANALYSES EFFECTUÉES

En ce qui concerne les méta-analyses de cartographie, si les interprétations sensorielles ou les configurations produits varient relativement peu en fonction de la méthode choisie, la construction des nouvelles ellipses de confiance (en accord avec le modèle Multi-MAM) permet d'obtenir beaucoup plus de discrimination entre les paires de produits.

Les performances des panels et des panélistes ont été étudiées à l'aide de la table MAM-CAP et du MultiMAM (global et multivarié). Ainsi, il a été mis en évidence que les personnes de plus de soixante ans et que les fumeurs dispersent moins leurs notes que les autres. De plus, des résultats établis dans d'autres méta-analyses ont été vérifiés (Brockhoff et al., 2015; Pineau, 2006).

Enfin, les méthodologies mises en place ont permis de quantifier l'importance des répétitions dans les épreuves de profil sensoriel et ont montré la difficulté de trouver l'équivalence de l'ajout d'une répétition en termes de nombre de sujets supplémentaires. Les répétitions ne paraissent pas nécessaires pour obtenir des résultats valables. Cependant, elles peuvent permettre de rassurer l'animateur de panel sur ses données, et de vérifier la répétabilité du panel. Le nombre de sujets additionnel équivalent à une deuxième répétition pour un panel de dix sujets est de 7 à 11 sujets en moyenne.

CHAPITRE 5 :

SYNTHESE ET DISCUSSION

Ce chapitre de synthèse et de discussion se décomposera en 4 parties. Un rappel de la démarche suivie permettra tout d'abord de mettre en évidence l'intérêt de coupler une approche théorique à une application sur base de données. Les résultats obtenus seront ensuite synthétisés, mis en perspective du contexte bibliographique présenté dans le chapitre 1 et discutés. Enfin, des recommandations à destination des animateurs de panel insisteront sur les implications pratiques de notre travail.

A. RAPPEL DE LA DEMARCHE SUIVIE

L'objectif de ce travail de thèse était de tester des méthodes alternatives pour l'analyse statistique de données de profil sensoriel, puis de valider ces méthodes par méta-analyses afin d'enrichir nos connaissances sur :

- les différences entre les méthodes statistiques de référence et les méthodes alternatives,
- l'impact de certaines caractéristiques de l'étude (âge des sujets, modalité sensorielle du descripteur...) sur les résultats observés.

La Figure 31 résume la démarche suivie pour répondre à ces objectifs, les différentes questions posées, et les liens avec les différentes méta-analyses effectuées.

Un travail bibliographique sur le profil sensoriel QDA a tout d'abord été mené et a confirmé que les trois principales analyses statistiques effectuées sur des données de profil sensoriel étaient **l'ANOVA à deux facteurs** (pour la détection de différences entre produits et le contrôle des tests de performances du panel et des sujets), les **tests post-hoc** (pour la constitution de groupes de produits) et **l'ACP** (pour la cartographie des produits). Nous nous sommes intéressés à l'ANOVA et à l'ACP et avons proposé des améliorations des calculs de performances et des techniques de cartographie en nous appuyant systématiquement sur des méta-analyses.

I. AMELIORATION DES CALCULS DE PERFORMANCES

L'ANOVA à deux facteurs ne prend pas en compte l'effet de scaling (le fait qu'un panéliste disperse plus ou moins ses notes par rapport au panel). Cet effet de scaling constitue un biais psychologique évoqué dans la littérature (Meilgaard et al., 1999). Pour prendre cet effet en compte dans l'analyse des données de profil sensoriel, Brockhoff et al. (2015) ont proposé le Mixed Assessor Model (MAM) qui inclut un coefficient de scaling permettant de détecter les sur- (ou sous-) dispersions des notes des sujets. Ce modèle a été utilisé durant cette thèse pour mettre en place des tests de performances unidimensionnelles du panel et des sujets. La table MAM-CAP permet de synthétiser les résultats de ces tests (Peltier et al., 2014) en utilisant des codes couleurs intuitifs. L'utilisation de cette table MAM-CAP sur différents jeux de données a montré qu'un même sujet pouvait avoir des coefficients de scaling très différents en fonction du descripteur noté. En s'appuyant sur le fait que l'erreur

de scaling était un biais psychologique, et par conséquent théoriquement commun à l'ensemble des descripteurs, un coefficient de scaling commun à l'ensemble des descripteurs a été proposé. Cette réflexion a abouti à la décomposition du coefficient de scaling du MAM en deux composantes : le scaling global (représentant une tendance générale à sur ou sous noter) et le scaling corrigé (représentant une hyper ou hypo sensibilité au descripteur) (Peltier et al., 2015d). Ces outils nous ont permis de mener des méta-analyses pour caractériser les performances du panel et des sujets. Le MAM a ensuite été étendu au multidimensionnel avec le MultiMAM (Peltier et al., 2015e) afin de traiter les performances multidimensionnelles. Cependant, une méta-analyse (Chapitre 4, A. III.) a montré que les tests de performance multidimensionnels ne permettaient pas d'améliorer la discrimination des produits à cause d'une hypersensibilité de la MANOVA à la réduction du nombre de degrés de liberté (résultant des calculs de coefficients de scaling).

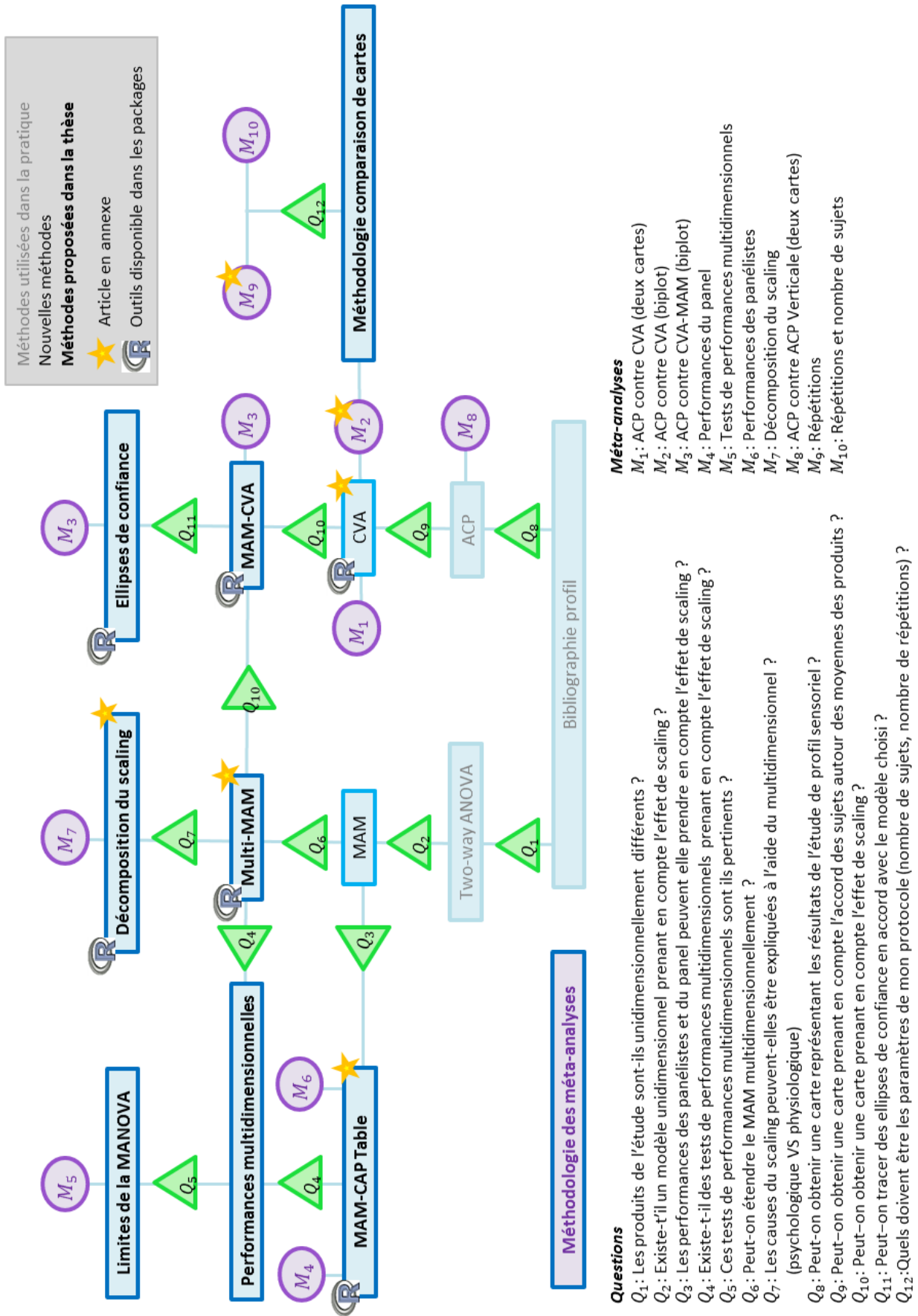


FIGURE 31 : SYNTHÈSE DU TRAVAIL EFFECTUÉ

II. AMELIORATION DES TECHNIQUES DE CARTOGRAPHIES DES PRODUITS

L'ACP, basée sur le tableau des moyennes des produits ne permet pas de prendre en compte la variabilité des sujets entre les moyennes des produits. Cela nous a conduit à mettre en avant une autre cartographie : la CVA, introduite par Schlich (2004) en analyse sensorielle. Cette méthode se base sur un modèle statistique afin de représenter non pas la variabilité des moyennes des produits, mais leur discrimination. Afin de comparer l'ACP et la CVA sur un grand nombre de jeux de données, une méthode de comparaison de cartes (Peltier et al., 2015b) a été mise en place. Cette méthodologie a permis de comparer également l'ACP et l'ACP Verticale.

Une carte basée sur le Multi-MAM, la CVA-MAM (Peltier et al., 2015e) a ensuite été proposée : elle permet de prendre en compte l'éventuel effet de scaling (global ou multivarié) des sujets dans la discrimination. Les tests multivariés comparant les produits deux à deux (T2 de Hotelling, visuellement représentés par les intersections des ellipses de confiance) ont également été adaptés au modèle (avec une prise en compte de l'effet sujet et de l'effet de scaling dans la détection de différences multidimensionnelles). Une méthode pour projeter uniquement le désaccord pur des sujets sur la carte (sans l'effet sujet, ni l'effet de scaling) a permis d'aboutir à des ellipses de confiance substantiellement plus petites.

III. UTILISATION DES METHODOLOGIES MISES EN PLACE POUR DES RESULTATS

D'ANALYSE SENSORIELLE

La mise en place de la méthodologie de comparaison de cartes a ensuite permis de revisiter un thème récurrent en analyse sensorielle : le nombre de répétitions nécessaire dans une épreuve de profil sensoriel. Le travail initié par Mammasse (2012) a été repris, puis complété par nos propres indicateurs de comparaison (y compris les indicateurs de comparaison de cartes) sur un plus grand nombre de jeux de données (Peltier et al., 2015c). Nous avons ensuite estimé une correspondance entre une répétition et un nombre de panélistes supplémentaires.

IV. ORIGINALITE DE LA DEMARCHE

Cette synthèse de la démarche de notre travail a bien montré l'utilité de traiter simultanément les avancées théoriques et leurs applications sur un grand nombre de jeux de données.

Cette approche permet de valider la pertinence de nouveaux modèles statistiques et de mettre en évidence de nouvelles problématiques. Par exemple, l'étude de la variabilité des coefficients des scaling a conduit à s'interroger sur la possibilité de les résumer en un unique indicateur de scaling, tous descripteurs confondus. De même, les méta-analyses ont montré les limites de l'extension multidimensionnelle des tests de performance unidimensionnels. La méta-analyse constitue donc un outil de validation des méthodes statistiques à utiliser, et ouvre des perspectives d'évolution des méthodes pertinentes.

Par ailleurs, les performances des sujets (notamment le scaling) en fonction de l'âge ou du sexe, mais aussi l'intérêt de répétitions supplémentaires ont pu être étudiés sur un grand nombre de jeux de données. Cette approche par méta-analyses peut mettre en évidence des résultats qu'il n'aurait pas été possible d'obtenir autrement.

Enfin, d'un point de vue pratique, pour tester les méthodes présentées dans cette thèse, nous avons dû les programmer nous-mêmes sous forme de bibliothèques R. Tester ces fonctions sur un grand nombre de jeux de données nous a permis de mettre en évidence des cas particuliers pour lesquels nos bibliothèques ne fonctionnaient pas. Les bibliothèques ont ainsi été améliorées afin de prendre en compte ces cas non prévus: la validation par méta-analyses a permis de rendre les bibliothèques R plus stables et plus robustes.

B. SYNTHÈSE ET DISCUSSION DES RESULTATS OBTENUS

Les résultats de cette thèse portent sur des méthodologies mises en place, des méthodes statistiques implémentées et de la documentation des pratiques du profil sensoriel.

I. RESULTATS METHODOLOGIQUES

Ces résultats comprennent les méthodologies de méta-analyses, de comparaisons de cartes et de comparaison de jeux de données.

1. METHODOLOGIES DE META-ANALYSES

i. SYNTHÈSE

La **méthodologie des méta-analyses** a été basée sur les études de Pineau (2006) et Mammasse (2012), mais aussi sur les méta-analyses de bases de données cliniques. La spécificité de notre approche consiste à proposer une forme de rapport originale en 5 étapes, permettant d'effectuer des méta-analyses en toute rigueur scientifique. La problématique de la méta-analyse est tout d'abord présentée. Des indicateurs statistiques sont ensuite mis en place pour y répondre. Connaissant les conditions d'application de ces indicateurs, des critères de sélection des jeux de données sont ensuite instaurés. Les jeux de données correspondant à ces critères sont ensuite sélectionnés, et vérifiés pour éviter de prendre en compte des **jeux de données atypiques** (nombre de produits ou de sujets trop grand ou trop petit...). A cette étape, on peut encore décider d'affiner les critères de sélection des jeux de données.

Ensuite, les critères sont calculés sur la sélection de jeux de données et leur distribution est tracée. Là encore, il est important de regarder **les distributions des indicateurs** (même si elle n'apparaît pas dans le rapport final) et de ne pas se restreindre à l'étude de la moyenne. Les distributions peuvent ainsi faire apparaître des groupes de jeux de données ou des jeux de données atypiques. Enfin, les distributions de ces critères nous permettent de conclure sur la problématique.

ii. DISCUSSION

Cette méthodologie est applicable à toute problématique et comprend les éléments présentés dans Mammasse (2012) et Pineau (2006). Ces étapes ont été ordonnées, de manière à ce que la méta-analyse soit optimisée (par exemple il est important d'établir les critères statistiques **avant** la sélection des jeux de données, puisque le calcul de certains critères peut impliquer des conditions particulières).

Le **choix des critères statistiques** comprend les éléments listés par Pineau (2006) : codage des données, filtrage des données, pondération des données, et choix du modèle statistique.

La **sélection des jeux de données** doit être effectuée dans l'esprit du QUORUM défini dans les méta-analyses médicales (Moher et al., 1999), c'est-à-dire qu'on ne retire pas un jeu de

données sans savoir pourquoi. En d'autres termes, il est aisé d'appliquer une fonction statistique à un ensemble de jeux de données, puis de « retirer » les jeux de données pour lesquels la fonction statistique ne donne pas de résultat. Cette habitude est à **proscrire** tant qu'on ignore pourquoi la fonction statistique ne donne pas de résultat pour ces jeux de données. Chacun des jeux de données problématiques doit faire l'objet d'une étude plus approfondie, permettant ou bien d'affiner les critères de sélection des jeux de données, ou bien d'améliorer la fonction statistique.

La **description de la base de données** par l'histogramme des nombres de produits, sujets, répétitions et descripteurs a été utilisée dans Mammasse (2012), Pineau (2006) et Brockhoff et al. (2015).

Cette méthodologie a été éprouvée par 10 méta-analyses dans le chapitre 4, pour lesquelles des difficultés de l'analyse des bases de données ont été mises en évidence (jeux de données incomplets, atypiques, non équilibrés, avec des données dupliquées, avec confusion entre session et répétition ou avec erreur de saisie). La **prise en compte de ces cas particuliers** (en améliorant les fonctions statistiques à appliquer ou en excluant les jeux de données concernés) **est primordiale pour la validité de la méta-analyse**.

Nous notons cependant que la dernière méta-analyse effectuée se basait, non pas sur un grand nombre de jeux de données (seulement 14 ont été utilisés) mais sur des techniques de ré-échantillonnage. Notre méthodologie n'a pas été réfléchie pour ce type d'analyse, et il pourrait y avoir des améliorations à fournir.

Dans le cadre d'une étude simultanée d'un grand nombre de jeux de données, nous pensons que **l'utilisation de cette méthodologie** (et du rapport associé) **permet d'attester de la qualité et de la rigueur scientifique de la méta-analyse effectuée**.

Il est à noter que l'approche de méta-analyses telle qu'elle est présentée dans cette thèse est à différencier du datamining classique (qui consiste à regrouper toutes les données et chercher des connections sans *a priori*).

2. METHODOLOGIE DE COMPARAISON DE CARTES

i. SYNTHÈSE

La **méthodologie de comparaison de cartes présentée dans Peltier et al. (2015b)** constitue une approche originale permettant de prendre en compte l'ensemble des éléments des cartes (configuration produit, interprétation sensorielle, groupes de produits, discrimination multidimensionnelle) pour les comparer. Cette méthodologie a été établie initialement pour comparer les cartes d'ACP et de CVA (cartes possédant un graphique représentant l'espace produit, et un cercle de corrélation pour les descripteurs). Elle a ensuite été adaptée au biplot.

ii. DISCUSSION

A notre connaissance, il n'existait pas de méthodologie permettant de comparer l'ensemble des caractéristiques des cartes. L'indicateur de comparaison des configurations des produits (coefficient RV) est souvent le critère unique utilisé pour comparer des cartes. Notre méthodologie a complété cette approche à l'aide de comparaisons des interprétations sensorielles de la carte, et des groupes de produits formés sur la carte (par T2 de Hotelling).

L'interprétation sensorielle est possible grâce à deux indicateurs : le maximum et la moyenne des angles entre les descripteurs pour une configuration produit fixée. La comparaison de CVA d'un jeu de données restreint à une répétition et de la CVA du jeu de données complet a mis en évidence que ces indicateurs ne permettent de comparer que des cartes avec les mêmes descripteurs. En d'autres termes, lorsque certains descripteurs sont significatifs pour un modèle et pas pour un autre, et que l'on trace les deux cartes associées, il n'est pas possible pour l'instant de comparer l'interprétation sensorielle des deux cartes (même si la plupart des descripteurs sont communs aux deux cartes). Une solution qui pourrait être mise en place serait de considérer uniquement les descripteurs communs aux deux cartes pour calculer ces indicateurs. Il faudrait néanmoins vérifier qu'il y a suffisamment de descripteurs communs aux deux cartes (et définir le « suffisamment »).

3. METHODOLOGIE DE COMPARAISON DE DEUX JEUX DE DONNEES DE PROFIL SENSORIEL

i. SYNTHÈSE

Une **méthodologie originale de comparaison** de deux jeux de données a également été mise en place. Elle porte sur les conclusions obtenues en termes de **discrimination (uni et multidimensionnelles)**, de **constitution de groupes des produits (uni et multidimensionnels)** et **d'accord (uni et multidimensionnel)** sur des jeux de données avec et sans répétitions. Les cartes de CVA obtenues sont également comparées à l'aide de **la méthodologie de comparaison de cartes**.

ii. DISCUSSION

La notion de proximité entre deux jeux de données n'est pas bien définie. Pour cela, il s'agirait de fixer des seuils de proximité pour chacun des indicateurs présentés (on pourrait fixer celui du coefficient RV à 0.9 par exemple), de manière arbitraire ou empirique.

Par ailleurs, certains des critères de comparaisons utilisés peuvent être biaisés par le nombre de sujets, de produits ou de descripteurs utilisés dans l'étude. Le coefficient RV par exemple est sensible au nombre d'observations comparées. Pour résoudre le problème, on peut avoir deux solutions : utiliser des critères standardisés ou vérifier les résultats par groupes de jeux de données (à nombre de produits faible, moyen ou élevé, par exemple...).

La méthodologie de comparaison de deux jeux de données de profil est basée sur le modèle à deux facteurs à effet sujet aléatoire. Ce modèle est utilisé pour la discrimination uni et multidimensionnelle, la constitution des groupes des produits et la comparaison de cartes.

Cette méthodologie devrait être également applicable pour un autre modèle statistique choisi (par exemple le MAM).

La méthodologie de comparaison de deux jeux de données de profil permet d'obtenir de multiples indicateurs de comparaison entre deux jeux de données. A partir d'une combinaison bien choisie de ces indicateurs, il pourrait être intéressant de concevoir un indicateur global que l'on considérerait comme une distance entre les jeux de données.

Cette méthodologie de comparaison de deux jeux de données peut s'appliquer au sein d'un même jeu de données (pour comparer des sous-jeux de données, comme ça a été le cas pour déterminer l'utilité des répétitions, ou le nombre de sujets supplémentaires équivalent à l'ajout d'une répétition) mais également entre jeux de données (par exemple issus de différents laboratoires, ou obtenus dans des pays différents). On pourrait également utiliser cette méthode pour déterminer le nombre de sujets optimal pour une séance de profil sensoriel.

II. RESULTATS STATISTIQUES (ET INFORMATIQUES)

Des améliorations des modèles statistiques habituellement utilisés pour analyser les données de profil sensoriel ont été proposées dans ce travail de thèse. Chacune des méthodes a été programmée dans une librairie R.

1. TESTS DE PERFORMANCES UNIDIMENSIONNELS ET TABLE MAM-CAP

i. SYNTHÈSE

Des tests de performances unidimensionnelles ont été mis en place à partir du Mixed Assessor Model. Ces tests permettent de prendre en compte l'effet de scaling (qui n'était pas pris en compte jusque-là). Des méta-analyses ont permis de démontrer que le MAM permettait d'améliorer la discrimination du panel et de réduire fortement son désaccord (Brockhoff et al. (2015)).

La table **MAM-CAP** (Peltier *et al.*, 2014) permet de visualiser les performances unidimensionnelles des panélistes et du panel, scaling compris. Cette table s'inspire de **l'interface de la table CAP** (Schlich, 1997b) pour visualiser les résultats de **tests de performances provenant du MAM**.

Le package MAMCAP associé permet d'obtenir la table CAP ou la table MAM-CAP. L'aide de ce package est disponible en annexe.

ii. DISCUSSION

La table **MAM-CAP pourrait être améliorée à l'aide du Multi-MAM global**. En effet, les coefficients de scaling utilisés dans la table CAP correspondent au Multi-MAM multivarié, qui comprend à la fois les causes psychologiques et physiologiques de l'effet de scaling. L'ajout d'une option à la table MAM-CAP permettant de représenter le scaling global et/ou le scaling corrigé (c'est-à-dire de mettre en évidence la décomposition du scaling pour chaque

sujet) pourrait fournir à l'animateur de panel un outil plus visuel que les sorties de la fonction MultiMAM.

Une autre amélioration envisagée est l'ajout d'une ligne multidimensionnelle à la table pour déterminer si les sujets (et le panel) ont été globalement performants (tous descripteurs confondus). Cet ajout impliquerait des tests de performances multidimensionnelles dont nous avons vu les limites. Des travaux ultérieurs pourront étudier comment contourner ces limites et proposer des indicateurs multidimensionnels.

2. CARTE ADAPTEE A L'ANALYSE SENSORIELLE : LA CVA

i. SYNTHÈSE

La **CVA** permet de construire une carte prenant en compte la variabilité des panélistes autour de la moyenne. Le biplot de la CVA, les ellipses de confiances associées au modèle choisi (sans projeter l'effet sujet) et les tables de Hotelling (permettant de tester multidimensionnellement les différences entre les produits) ont été programmés. Cette carte est disponible dans le package CVA-pack (aide disponible en annexe). L'application de la CVA dans l'analyse de profil sensoriel a été décrite dans Peltier et al. (2015a).

ii. DISCUSSION

En ce qui concerne les ellipses de confiance, il existe deux écoles : des ellipses construites par projection directe des notes obtenues par les sujets de l'étude (Saporta, 1986) et des ellipses basées sur du ré-échantillonnage (bootstrap) utilisées par Cadoret and Husson (2013). Ces ellipses n'ont pas les mêmes significations. Lorsque les premières sont disjointes, cela signifie que le panel utilisé a trouvé des différences multidimensionnelles entre les produits. Lorsque les secondes sont disjointes, cela signifie que si on avait sélectionné d'autres panélistes, on aurait probablement trouvé des différences multidimensionnelles entre les produits. Par conséquent, les ellipses de bootstrap sont généralement plus grandes que celle de projections simples.

La CVA est basée sur le calcul d'une inversion de matrice, laquelle n'est pas toujours possible. Nous avons observé lors de la méta-analyse que de tels problèmes étaient relativement rares avec les données d'analyse sensorielle.

La Partial Least Squares-Discriminant Analysis (PLS-DA) est une technique de cartographie utilisée par Rossini, Verdun, Cariou, Qannari, and Fogliatto (2012) pour produire des cartes de profil sensoriel. Cette carte présente l'avantage de ne pas être soumise aux problèmes de conditionnement, ou de non inversibilité des matrices. Cette carte constitue une alternative à la CVA, et il pourrait être envisageable de s'inspirer de la PLS-DA pour construire une PLS-CVA, basée sur un modèle statistique.

3. UN MODELE MULTIDIMENSIONNEL PRENANT LE SCALING EN COMPTE

i. SYNTHÈSE

Le **Multi-MAM** (Peltier *et al.*, 2015e) est le modèle multidimensionnel associé au MAM. Le package associé permet d'obtenir les coefficients de scaling global et corrigé, mais aussi la décomposition des notes et la **MAM-CVA** (carte associée au modèle Multi-MAM) avec construction d'ellipses de confiance en adéquation avec le modèle choisi.

Enfin, **des limites de la MANOVA** ont été mises en évidence lors de la généralisation des tests unidimensionnels à des tests multidimensionnels.

ii. DISCUSSION

Romano *et al.* (2015) ont proposé une extension du multiplicative Assessor Model. Cette extension est basée sur une approche par composantes principales et considère un coefficient de scaling par descripteur et sujet. Cela consiste à effectuer une ACP de la matrice constituée des coefficients de scaling MAM. Ce modèle s'intéresse exclusivement à l'effet de scaling.

Notre approche se différencie de ce modèle sur différents points :

- l'utilisation d'un coefficient de scaling unique par panéliste,
- la généralisation du MAM (et pas du multiplicative Assessor Model)
- l'utilisation de la CVA permettant de visualiser la discrimination des produits.

III. SYNTHÈSE ET DISCUSSION SUR LES RESULTATS D'ANALYSE SENSORIELLE

1. PRATIQUES D'ANALYSE SENSORIELLE MISES EN EVIDENCE PAR LA SENSOBASE

i. SYNTHÈSE

L'étude de la SensoBase a montré qu'un jeu de données médian comprenait 5 produits, 11 panélistes, 22 descripteurs et 2 répétitions. On a remarqué que les femmes sont globalement plus présentes que les hommes dans les études de profil sensoriel (70% de femmes). Les sujets sont âgés de 48 ans en moyenne et la moitié des sujets ont entre 40 et 60 ans.

Dans la SensoBase, on a remarqué un nombre de zéros élevé, particulièrement pour les descripteurs d'odeur. Par ailleurs, une méta-analyse a permis de montrer que considérer l'effet sujet comme fixe concluait à tort à des différences pour 23% des descripteurs.

ii. DISCUSSION

La composition des panels n'est pas souvent représentative de la population globale (panels majoritairement féminins et âgés entre 40 et 60 ans). Cela peut s'expliquer par le fait que ces sujets sont plus disponibles et peut être plus intéressés par l'analyse sensorielle.

Le nombre de zéros peut s'avérer problématique pour le respect des hypothèses de l'ANOVA. Il pourrait refléter un comportement de « absence/présence » du descripteur dans le produit. Ce type de descripteur pourrait ne pas être analysé comme un descripteur noté sur une échelle quantitative continue, mais comme une variable binaire.

Le choix du modèle aléatoire contre le modèle à effet fixes en profil sensoriel (Næs & Langsrud, 1998) a un réel impact sur les résultats de l'analyse des données de profil sensoriel, il est donc **primordial pour l'animateur de panel de choisir le modèle aléatoire**, pour ne pas conclure à tort à des discriminations entre ses produits, qui ne seraient pas robustes par rapport à l'échantillonnage des sujets.

2. PERFORMANCES EN PROFIL SENSORIEL

i. SYNTHÈSE

Aucune différence significative n'a été détectée entre les performances unidimensionnelles des hommes et des femmes. En revanche, une tendance montrant que les **fumeurs seraient moins performants** que les autres panélistes a été observée, mais cette tendance reste à confirmer avec des études spécifiques. Les résultats les plus significatifs ont été obtenus sur l'âge des panélistes : les **plus de soixante ans** (les plus âgés du panel) **ont tendance à utiliser une portion de l'échelle moindre** par rapport aux autres panélistes et par conséquent **à être plus répétables**. Ils apparaissent également **légèrement moins discriminants** que les autres panélistes.

L'utilisation du coefficient de scaling global a confirmé le fait que **les fumeurs et les plus de soixante ans ont tendance à moins disperser leurs notes** que les autres, **tous descripteurs confondus**. Étonnamment, une personne qui avait tendance à sur-disperser ses notes pour un jeu de données fixé, n'aura pas forcément tendance à les sur-disperser pour un autre jeu de données. Enfin, l'analyse du scaling corrigé des sujets par modalité sensorielle a montré des **hyper ou hypo sensibilités plus présentes pour les descripteurs d'odeurs et d'arômes que pour les autres descripteurs**. Par ailleurs, si, pour un sujet, un coefficient de scaling négatif est obtenu, cela montre que ce sujet a ordonné les produits à l'inverse du panel, et indique ainsi un désaccord avec le panel plutôt qu'un effet d'échelle : l'effet de scaling ne devrait pas être corrigé. Ces **problèmes de négativité ne se produisent que très rarement avec l'utilisation du scaling global** (moins de 1%).

ii. DISCUSSION

Les résultats obtenus sont comparables à ceux de Pineau (2006) (même si il n'avait pas utilisé les mêmes classes d'âge). De plus, ils confirment des hypothèses présentées dans la bibliographie en ce qui concerne les descripteurs de goût et d'odeur. (Ahlström et al., 1987; Booth et al., 1982; Krut, 1961; Rawson et al., 2012).

Étant donné qu'on a observé peu de différences entre les performances des hommes et des femmes, on peut supposer que l'équilibrage homme/femme n'est pas primordial dans une

étude de profil sensoriel. En revanche, n'utiliser que des fumeurs ou que des personnes âgées de plus de 60 ans pourrait conduire à des panels légèrement moins discriminants.

La décomposition du scaling en deux composantes (psychologique et physiologique) rejoint les étapes successives de la perception du goût et de sa notation présenté dans Lim (2011). De plus, il permet de prendre en compte un biais présenté dans Meilgaard et al. (1999) : celui d'avoir un biais dû au caractère timide (qui sous-disperse ses notes) ou capricieux (qui sur-disperse ses notes).

Ces travaux ont ouvert la porte à d'autres expérimentations spécifiques pour vérifier certains des résultats obtenus par méta-analyses. Par exemple, nos résultats sur les fumeurs pourraient être illustrés avec des jeux de données et des recrutements spécifiques afin de vérifier nos conclusions et de démontrer l'effet du fait de fumer, ou d'arrêter de fumer, sur la sensibilité physiologique des panélistes. Par ailleurs, les causes de la variation du scaling global en fonction du jeu de données pourraient être étudiées.

3. CARTES EN PROFIL SENSORIEL

i. SYNTHÈSE

La comparaison des différentes méthodes de cartographie a montré une **stabilité globale des configurations des produits et des interprétations sensorielles des cartes**. En revanche, la taille des ellipses de confiance varie selon la représentation choisie, ainsi que le diagnostic de groupes de produits (deux produits sont dans le même groupe si leurs ellipses se chevauchent).

ii. DISCUSSION

Ces résultats peuvent permettre de rassurer la communauté de l'analyse sensorielle, habituée aux ACP. Utiliser la CVA à la place de l'ACP ne conduit pas à des différences majeures d'interprétations sensorielles des cartes. Cependant, des petites différences existent et il est théoriquement plus juste d'utiliser la CVA que l'ACP avec des données de profil sensoriel, ne serait-ce que pour avoir une carte cohérente avec le modèle unidimensionnel utilisé.

Par la suite, il pourrait également être possible d'adapter la méthodologie utilisée pour la comparaison de cartes d'ACP et de CVA aux cartes de CVA et de PLS-DA.

IV. DISCUSSION GENERALE, LIMITES ET PERSPECTIVES

Le Tableau 25 représente l'ensemble des analyses rendues disponibles par ce travail, en fonction de l'approche adoptée (prise en compte ou non du scaling).

Scaling	Modele unidimensionnel	Performances unidimensionnelles	Groupes unidimensionnels	Modèle multidimensionnel	Cartes
Non corrigé	Modèle à 2 facteurs à effet sujet aléatoire	Table CAP	HSD.test (avec terme d'erreur = interaction produit*sujet)	Modèle multidimensionnel à effet sujet aléatoire	CVA (à deux facteurs)
Corrigé systématiquement (scaling multivarié)	MAM	Table MAM-CAP	Pas de méthode spécifique	MultiMAM (multivarié)	MAM-CVA (multivarié)
Corrigé uniquement lorsqu'il est psychologique (scaling global)	-	Coefficients de scaling globaux et corrigés	Pas de méthode spécifique	MultiMAM (global)	MAM-CVA (globale)

TABEAU 25 : RECAPITULATIF DES ANALYSES ASSOCIEES AUX DIFFERENTES APPROCHES

On remarque que ce travail de thèse ne traite pas des tests de groupes de produits effectués en prenant le scaling en compte. Usuellement, ce sont des tests LSD ou de Tukey qui sont réalisés. Ces tests ne prennent pas en compte l'effet de scaling. Les tests LSD ou de Tukey devraient être utilisés avec le terme d'erreur adapté à la correction de scaling (le désaccord pur).

Par ailleurs, pour l'instant on ne peut pas traiter les descripteurs un à un en corrigeant uniquement le scaling global. En d'autres termes, les tests de discrimination unidimensionnels ne prennent pas le scaling global en compte.

Un travail d'homogénéisation de programmation pourrait être fait pour aboutir à une librairie R permettant d'analyser complètement les données de profil sensoriel suivant un modèle fixé. Cette librairie R pourra être utilisée et mise à disposition dans TimeSens (cf Chapitre 1. B. IV.), pour être disponible en « presse-bouton ».

A l'issue de ce travail de thèse, il reste certains points d'analyse des données de profil sensoriel qui mériteraient d'être étudiés.

Les librairies R fonctionnent bien dans le cas de jeux de données complets et équilibrés. En revanche, il reste à les **adapter pour les jeux de données non équilibrés ou incomplets**. Dans le cas où seules quelques données manquantes sont présentes, ces données peuvent être imputées. Nos packages R devraient cependant être retravaillés pour prendre en compte ces cas particuliers.

L'importance de séparer les analyses par modalité sensorielle des descripteurs a été démontrée, tant pour les performances que pour les représentations graphiques. D'une part, le panel a tendance à être plus **performant pour les descripteurs visuels ou de texture** que pour les descripteurs d'arômes et d'odeur. D'autre part, les cartes d'ACP et de CVA sont davantage similaires lorsque l'on traite les modalités sensorielles indépendamment. Le calcul d'un coefficient de scaling global par modalité sensorielle permettrait de détecter certains sujets plus ou moins sensibles aux odeurs par exemple. Ainsi, des profils de sensibilité à la

vision, au goût etc... pourraient être mis en évidence, et les sujets classifiés en fonction de leur profil : sujets hypersensibles à la texture, aux odeurs, au visuel...

C. RECOMMANDATIONS A L'ISSUE DE CE TRAVAIL

Cette thèse a abouti à des recommandations sur le protocole à appliquer lors d'une tâche de profil sensoriel et sur les analyses statistiques des données à mener.

I. RECOMMANDATIONS SUR LE PROTOCOLE DE PROFIL SENSORIEL

Nos travaux ont montré que l'animateur de panel doit garder à l'esprit que **l'âge et le fait de fumer ou non influent sur les performances des panélistes**. Il semble que les panélistes fumeurs et les panélistes plus âgés soient moins sensibles que les autres.

Le **nombre de sujets ou de répétitions** est souvent limité par les coûts financiers. De plus, l'augmentation du nombre de sujets implique souvent des problèmes matériels tels que la disponibilité d'un nombre suffisant de cabines de dégustations, ou les problèmes de recrutement. Le travail effectué dans le chapitre 4 (partie C) montre que les répétitions apportent une quantité d'informations assez peu élevée en moyenne et que l'équivalence d'une deuxième répétition en nombre de sujets est difficile à estimer. Nous recommandons cependant **l'utilisation d'une deuxième répétition** qui permet d'évaluer la répétabilité du panel, et constitue une sécurité pour l'animateur de panel, ainsi que le recrutement de **10 sujets au moins** d'après l'étude de Mammasse (2012).

II. RECOMMANDATIONS SUR L'ANALYSE DES DONNEES DE PROFIL SENSORIEL

1. RECOMMANDATIONS POUR L'ANALYSE DES PERFORMANCES EN PROFIL SENSORIEL

Le modèle de référence de l'analyse d'un profil sensoriel est le modèle à deux facteurs avec effet sujet aléatoire. La méta-analyse présentée en guise d'exemple dans le Chapitre 2 a permis de montrer l'importance d'utiliser ce modèle plutôt que celui à effets fixes. Cependant, aucun de ces modèles ne prend en compte l'effet de scaling.

Lors de la phase d'entraînement des panélistes, nous conseillons **d'utiliser la table MAM-CAP** afin de déterminer si le panel et les panélistes sont performants sur l'ensemble des descripteurs pour éventuellement les ré-entraîner si ce n'est pas le cas. Pour cela, le package MAM-CAP peut être utilisé avec les corrélations de Pearson pour les tests d'accord. Pour obtenir les coefficients de scaling globaux et corrigés, le package MultiMAM peut être utilisé.

La table MAM-CAP peut également être construite par modalité sensorielle afin de détecter d'éventuelles défaillances de sujets par rapport à une modalité sensorielle spécifique.

2. RECOMMANDATIONS POUR LA CONSTRUCTION DE CARTES PRODUITS EN PROFIL SENSORIEL

Lors de la phase de mesure, la **MAM-CVA** (avec scaling global) devrait être effectuée dans le but de représenter l'expérimentation dans sa globalité. Les similarités multidimensionnelles

entre produits seront visualisées par la présence de traits reliant deux produits non significativement différents dans la carte.

Lorsque c'est possible, nous recommandons **d'effectuer les cartes par modalité sensorielle**. Dans le cas où les configurations produits seraient similaires pour l'ensemble des modalités sensorielles (et dans ce cas seulement), la carte tous descripteurs confondus pourra être construite.

3. SYNTHÈSE DES RECOMMANDATIONS À DESTINATION DE L'ANIMATEUR DE PANEL

Les recommandations à destination de l'animateur de panel sont synthétisées dans le Tableau 26 et le Tableau 27.

Questions	Recommandations
Nombre de répétitions ?	2 au plus
Nombre de sujets ?	10 ou plus
Impact du genre sur les performances ?	Non. L'équilibre homme femme n'est pas nécessaire dans le panel.
Impact du tabagisme sur les performances ?	Oui. Les fumeurs sont moins discriminants que les autres.
Impact de l'âge sur les performances	Oui. Les personnes plus âgées semblent être moins discriminantes que les autres.

TABLEAU 26 : RESUME DES RECOMMANDATIONS AU NIVEAU DU PROTOCOLE DE PROFIL SENSORIEL

Questions	Recommandations		
	<i>Sans prendre le scaling en compte</i>	<i>En prenant en compte le scaling du MAM</i>	<i>En prenant en compte uniquement le scaling global</i>
Parti pris			
Quelle carte ?	CVA	MAM-CVA multivariée	MAM-CVA globale
Quels tests de performances	Table CAP	Table MAM-CAP	Analyse du scaling global et corrigé en plus de la table CAP et MAM-CAP
Analyse par modalité sensorielle ?	Oui. Pour les cartes et les performances, cela peut apporter des informations supplémentaires.		

TABLEAU 27 : RESUME DES RECOMMANDATIONS AU NIVEAU DES ANALYSE DE DONNEES DE PROFIL SENSORIEL

Concernant le nombre de sujets, on notera que la recommandation peut varier en fonction de la modalité sensorielle du descripteur. Enfin, une seconde répétition peut constituer une

sécurité pour l'animateur de panel et permettre de vérifier la répétabilité de ses sujets, alors qu'une troisième répétition n'apporte pas suffisamment d'informations supplémentaires.

CONCLUSION

Notre travail visait à revisiter les analyses statistiques des données de profil sensoriel à l'aide d'une approche base de données. La SensoBase, contenant plus de mille jeux de données de profil sensoriel a été une ressource indispensable à la réalisation de cet objectif. L'utilisation d'une méthodologie de méta-analyse sur la SensoBase a permis d'obtenir différents résultats utiles à l'animateur de panel de profil sensoriel. On a notamment montré que les fumeurs et les personnes âgées de plus de soixante ans avaient tendance à moins disperser leurs notes que les autres panélistes et à moins discriminer les produits. En revanche, aucun impact du genre sur les performances n'a été observé. Par ailleurs, de nouvelles méthodes statistiques ont été proposées pour traiter les performances des sujets et la cartographie des produits. Ainsi, l'effet du scaling (tendance à sur- ou sous-disperser ses notes par rapport au reste du panel) a été étudié et intégré dans des tests de performances adaptés (MAM-CAP table). Il a ensuite été décomposé en une composante psychologique (scaling global) et physiologique (scaling corrigé). Les techniques de cartographie (ACP et CVA) ont également été améliorées par des ellipses de confiance en accord avec le modèle choisi et la prise en compte du scaling dans la conception de la carte (MAM-CVA). Des outils ont été mis à disposition des animateurs de panel pour appliquer ces recommandations. Ainsi, les bibliothèques MAMCAP, CVAPack et MultiMAM permettent d'obtenir respectivement la table de performance des sujets, la carte CVA, et les résultats sur le scaling global (MAM-CVA et coefficient de scaling global). Ces bibliothèques ont été également intégrées dans le logiciel TimeSens pour une interface plus « user-friendly ». Finalement, les différences entre les méthodes statistiques (MAM, CVA, CVA-MAM) proposées dans ce manuscrit et les méthodes utilisées actuellement (ACP, ANOVA avec modèle à deux facteurs et effet sujet aléatoire) ont été quantifiées par méta-analyses à l'aide de la SensoBase.

La SensoBase contenant uniquement des études concernant des produits alimentaires, on peut noter que ce travail est également applicable pour d'autres types de produits (issus des cosmétiques, de l'industrie automobile...). Des résultats différents (et des recommandations différentes) spécifiques au type de produit évalué pourraient ainsi être obtenus. En cela, nous encourageons vivement les détenteurs de telles bases de données à effectuer les méta-analyses proposées dans cette thèse. Cela pourra être rendu possible ultérieurement par l'implémentation des méta-analyses de cette thèse dans le logiciel TimeSens.

Par ailleurs, les méthodes statistiques proposées et validées par méta-analyses dans cette thèse pourraient être utilisables dans d'autres domaines scientifiques utilisant l'être humain comme instrument de mesure (psychologie, économie, éventuellement médecine...). En effet, la problématique du scaling n'est pas spécifique à l'analyse sensorielle, mais se généralise à l'ensemble des disciplines où des êtres humains sont amenés à attribuer des notes à différents descripteurs. La prise en compte statistique des composantes psychologiques et physiologiques du scaling de ces sujets pourrait apporter des éclairages nouveaux à des questions soulevées dans ces domaines.

Pour terminer, aujourd'hui, tous les domaines sont amenés à stocker des données (stockage facilité par les progrès informatiques), et les détenteurs de ces bases de données sont confrontés à un problème double : d'une part, trouver la méthode la plus adaptée pour analyser leurs jeux de données individuellement ; et d'autre part, exploiter au mieux la base de données formée par l'ensemble des jeux de données pour en extraire de la connaissance. Ce travail de thèse a établi une démarche scientifique qui a permis de répondre à ces questions dans le cadre spécifique de l'analyse du profil sensoriel. Il a mis en évidence l'intérêt de traiter conjointement les avancées statistiques méthodologiques et leur traitement par méta-analyses sur un grand nombre de jeux de données. Les avancées statistiques permettent d'extraire de nouvelles informations des bases de données et l'utilisation des bases de données permet de valider, vérifier et améliorer la programmation des avancées statistiques par la confrontation à la pratique. Cette approche originale constitue donc une stratégie « gagnant-gagnant », dont les apports peuvent être extrêmement riches. C'est pourquoi cette démarche pourrait (et devrait) être étendue à tout domaine stockant et analysant des jeux de données.

BIBLIOGRAPHIE

- Ahlström, R., Berglund, B., Berglund, U., Engen, T., & Lindvall, T. (1987). A comparison of odor perception in smokers, nonsmokers, and passive smokers. *American Journal of Otolaryngology*, 8(1), 1-6.
- Amoa-Awua, W. K., Awusi, B., Owusu, M., Appiah, V., Ofori, H., Thorsen, L., et al. (2014). Reducing the atypical odour of dawadawa: Effect of modification of fermentation conditions and post-fermentation treatment on the development of the atypical odour of dawadawa. *Food Control*, 42(o), 335-342.
- Aquino, L. F. M. C., Silva, A. C. O., Freitas, M. Q., Felicio, T. L., Cruz, A. G., & Conte-Junior, C. A. (2014). Identifying cheese whey an adulterant in milk: Limited contribution of a sensometric approach. *Food Research International*, 62(o), 233-237.
- Ares, G., & Jaeger, S. R. (2015). 11 - Check-all-that-apply (CATA) questions with consumers in practice: experimental considerations and impact on outcome. In J. Delarue, J. B. Lawlor & M. Rogeaux, *Rapid Sensory Profiling Techniques*: Woodhead Publishing.
- Bagdi, A., Szabó, F., Gere, A., Kókai, Z., Sipos, L., & Tömösközi, S. (2014). Effect of aleurone-rich flour on composition, cooking, textural, and sensory properties of pasta. *LWT - Food Science and Technology*, 59(2, Part 1), 996-1002.
- Ballay, S., Loescher, E., & Gazano, G. (2015). 19 - Projective Flash Profile from experts to consumers: a way to reveal fragrance language. In J. Delarue, J. B. Lawlor & M. Rogeaux, *Rapid Sensory Profiling Techniques*: Woodhead Publishing.
- Bechhoff, A., Cissé, M., Fliedel, G., Declémy, A.-L., Ayessou, N., Akissoe, N., et al. (2014). Relationships between anthocyanins and other compounds and sensory acceptability of Hibiscus drinks. *Food Chemistry*, 148(o), 112-119.
- Biasoto, A. C. T., Netto, F. M., Marques, E. J. N., & da Silva, M. A. A. P. (2014). Acceptability and preference drivers of red wines produced from *Vitis labrusca* and hybrid grapes. *Food Research International*, 62(o), 456-466.
- Booth, P., Kohrs, M. B., & Kamath, S. (1982). Taste acuity and aging: A review. *Nutrition Research*, 2(1), 95-109.
- Brandt, M. A., Skinner, E., & Coleman, J. (1963). Texture profile Method. *Journal of Food Science*, 28, 404-410.
- Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. *Food Quality and Preference*, 39(o), 156-166.

- Cadena, R. S., Caimi, D., Jaunarena, I., Lorenzo, I., Vidal, L., Ares, G., et al. (2014). Comparison of rapid sensory characterization methodologies for the development of functional yogurts. *Food Research International*, 64(0), 446-455.
- Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference*, 28(1), 106-115.
- Cairncross, W. E., & Sjöström, L. B. (1950). Flavor profil - a new approach to flavor problems. *Food Technology*, 4, 308-311.
- Canto, A. C. V. C. S., Lima, B. R. C. C., Suman, S. P., Lazaro, C. A., Monteiro, M. L. G., Conte-Junior, C. A., et al. (2014). Physico-chemical and sensory attributes of low-sodium restructured caiman steaks containing microbial transglutaminase and salt replacers. *Meat Science*, 96(1), 623-632.
- Castro-Vázquez, L., Leon-Ruiz, V., Alañon, M. E., Pérez-Coello, M. S., & González-Porto, A. V. (2014). Floral origin markers for authenticating Lavandin honey (*Lavandula angustifolia* x *latifolia*). Discrimination from Lavender honey (*Lavandula latifolia*). *Food Control*, 37(0), 362-370.
- Chojnicka-Paszun, A., Doussinault, S., & de Jongh, H. H. J. (2014). Sensorial analysis of polysaccharide-gelled protein particle dispersions in relation to lubrication and viscosity properties. *Food Research International*, 56(0), 199-210.
- Cucherat, M. (1997). Manuel pratique de méta-analyse des essais thérapeutiques. In. Lyon.
- Czarnacka-Szymani, J., & Jezewska-Zychowicz, M. Impact of nutritional information on consumers' acceptance of cheese with reduced sodium chloride content. *International Dairy Journal*(0).

- Delarue, J. (2015). 6 - Flash Profile, its evolution and uses in sensory and consumer science. In J. Delarue, J. B. Lawlor & M. Rogeaux, *Rapid Sensory Profiling Techniques*: Woodhead Publishing.
- Dettmar, B., Plattner, M., & Vetter, C. (2014). Underdetermination in Multivariate data analysis. In, *13th symposium on statistical methods for the food industry*.
- Devezeaux de Lavergne, M., van Delft, M., van de Velde, F., van Boekel, M. A. J. S., & Stieger, M. Dynamic texture perception and oral processing of semi-solid food gels: Part 1: Comparison between QDA, progressive profiling and TDS. *Food Hydrocolloids*(0).

- Egea, M., Linares, M. B., Garrido, M. D., Villodre, C., Madrid, J., Orengo, J., et al. (2014). Crude glycerine inclusion in Limousin bull diets: Animal performance, carcass characteristics and meat quality. *Meat Science*, 98(4), 673-678.
- Egger, M., & Smith, G. D. (1997). Meta-analysis. Potentials and promise. *British Medical Journal*, 315, 1371-1374.

- Escoufier, Y. (1973). Le Traitement des Variables Vectorielles. *Biometrics*, 29(4), 751-760.
- Esposito, G., Masucci, F., Napolitano, F., Braghieri, A., Romano, R., Manzo, N., et al. (2014). Fatty acid and sensory profiles of Caciocavallo cheese as affected by management system. *Journal of Dairy Science*, 97(4), 1918-1928.

- F**eng, Y., Cai, Y., Su, G., Zhao, H., Wang, C., & Zhao, M. (2014). Evaluation of aroma differences between high-salt liquid-state fermentation and low-salt solid-state fermentation soy sauces from China. *Food Chemistry*, 145(0), 126-134.
- Feng, Y., Chen, Z., Liu, N., Zhao, H., Cui, C., & Zhao, M. (2014). Changes in fatty acid composition and lipid profile during koji fermentation and their relationships with soy sauce flavour. *Food Chemistry*, 158(0), 438-444.
- Francis, F., Talhouk, S., Batal, M., & Olabi, A. (2014). Sensory and quality parameters of raw and processed Chicory-Hindbeh, a commonly consumed dark leafy green in Lebanon (*Cichorium intybus* L.) during frozen storage. *LWT - Food Science and Technology*, 58(1), 230-238.

- G**abriel, K., R (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467.
- Genisheva, Z., Vilanova, M., Mussatto, S. I., Teixeira, J. A., & Oliveira, J. M. (2014). Consecutive alcoholic fermentations of white grape musts with yeasts immobilized on grape skins - Effect of biocatalyst storage and SO₂ concentration on wine characteristics. *LWT - Food Science and Technology*, 59(2, Part 1), 1114-1122.
- Golub, G. H., & Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics*, 2(2), 205-224.
- Gomes, C. L., Pflanzner, S. B., Cruz, A. G., de Felício, P. E., & Bolini, H. M. A. (2014). Sensory descriptive profiling and consumer preferences of beef strip loin steaks. *Food Research International*, 59(0), 76-84.
- Gower, J. (1975). Generalized Procrustes Analysis. *Psychometrika*, 40(1).
- Gower, J., Lubbe, S., & Le Roux, N. (2011). Understanding biplot. In L. John Wiley and Sons.

- K**rut, L. H. P., M.; J., Bronte-Stewart, M.D.,. (1961). Taste perception in smokers and non-smokers. *British Medical Journal*, 1:384.

- H**ixson, S. M., Parrish, C. C., & Anderson, D. M. (2014). Full substitution of fish oil with camelina (*Camelina sativa*) oil, with partial substitution of fish meal with

camelina meal, in diets for farmed Atlantic salmon (*Salmo salar*) and its effect on tissue lipids and sensory quality. *Food Chemistry*, 157(0), 51-61.

- L**aguna, L., Primo-Martín, C., Varela, P., Salvador, A., & Sanz, T. (2014). HPMC and inulin as fat replacers in biscuits: Sensory and instrumental evaluation. *LWT - Food Science and Technology*, 56(2), 494-501.
- Larson-Powers, N., & Pangorn, R. M. (1978). Paired comparison and time-intensity measurement of the sensory properties of beverages and gelatins containing sucrose or synthetic sweeteners. *Journal of Food Science*, 43(41).
- Latocha, P., Krupa, T., Jankowski, P., & Radzanowska, J. (2014). Changes in postharvest physicochemical and sensory characteristics of hardy kiwifruit (*Actinidia arguta* and its hybrid) after cold storage under normal versus controlled atmosphere. *Postharvest Biology and Technology*, 88(0), 21-33.
- Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference*, 22(8), 733-747.
- Lorentzen, G., Skuland, A. V., Sone, I., Johansen, J.-O., & Rotabakk, B. T. (2014). Determination of the shelf life of cluster of the red king crab (*Paralithodes camtschaticus*) during chilled storage. *Food Control*, 42(0), 207-213.
- Lorido, L., Estévez, M., & Ventanas, S. (2014). A novel approach to assess temporal sensory perception of muscle foods: Application of a time-intensity technique to diverse Iberian meat products. *Meat Science*, 96(1), 385-393.
- Lü, Y.-g., Chen, J., Li, X.-q., Ren, L., He, Y.-q., & Qu, L.-B. (2014). Study on processing and quality improvement of frozen noodles. *LWT - Food Science and Technology*, 59(1), 403-410.

- M**ammasse, N. (2012). *Le nombre de sujets dans les panels d'analyse sensorielle : une approche base de données*. Université de Bourgogne, Dijon.
- Mardia, K. V., Kent, J. T., & Bibbly, J. M. (1989). *Multivariate Analysis*.
- Meilgaard, M., Gail Vance Civille, B. S., & Carr, T. B. (1999). *Sensory evaluation techniques (3. ed.)*. USA: Taylor & Francis Group LLC.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet*, 354(9193), 1896-1900.
- Morais, E. C., Cruz, A. G., Faria, J. A. F., & Bolini, H. M. A. (2014). Prebiotic gluten-free bread: Sensory profiling and drivers of liking. *LWT - Food Science and Technology*, 55(1), 248-254.
- Muñoz, A. M., & Civille, G. V. (1992). The Spectrum descriptive analysis method. In Am., *ASTM Manual Series MNL 13, Manual On Descriptive Analysis Testing*.

- N**æs, T., & Langsrud, Ø. (1998). Fixed or random assessors in sensory profiling? *Food Quality and Preference*, 9(3), 145-152.

Nony, P., Boissel, J. P., Lièvre, M., Cucherat, M., Haugh, M. C., & Dayoub, G. (1995). Introduction à la méthodologie métaanalytique. *La Revue de Médecine Interne*, 16(7), 536-546.

- P**aulsen, M. T., Nys, A., Kvarberg, R., & Hersleth, M. (2014). Effects of NaCl substitution on the sensory properties of sausages: Temporal aspects. *Meat Science*, 98(2), 164-170.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243-1246.
- Peltier, C., Brockhoff, P. B., Visalli, M., & Schlich, P. (2014). The MAM-CAP table: A new tool for monitoring panel performances. *Food Quality and Preference*, 32, Part A(o), 24-27.
- Peltier, C., Visalli, M., & Schlich, P. (2015a). Canonical Variate Analysis of sensory profiling data. *Journal of Sensory Studies*, (accepted).
- Peltier, C., Visalli, M., & Schlich, P. (2015b). Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies. *Food Quality and Preference*, 40, Part B(o), 326-333.
- Peltier, C., Visalli, M., & Schlich, P. (2015c). Do we need to replicate in profile sensory studies. *Food Quality and Preference*, (to be submitted).
- Peltier, C., Visalli, M., & Schlich, P. (2015d). Multiplicative decomposition of the scaling effect in the Mixed Assessor Model into a descriptor specific and an overall coefficients. *Food Quality and Preference*, (submitted).
- Peltier, C., Visalli, M., & Schlich, P. (2015e). Taking the scaling effect into account in multidimensional analysis with MultiMAM. *Journal of Statistical Software*, (to be submitted).
- Picouet, P. A., Sárraga, C., Cofán, S., Belletti, N., & Dolors Guàrdia, M. Effects of thermal and high-pressure treatments on the carotene content, microbiological safety and sensory properties of acidified and of non-acidified carrot juice. *LWT - Food Science and Technology*(o).
- Pimentel, T. C., Madrona, G. S., & Prudencio, S. H. Probiotic clarified apple juice with oligofructose or sucralose as sugar substitutes: Sensory profile and acceptability. *LWT - Food Science and Technology*(o).
- Pineau, N. (2006). *La performance en analyse sensorielle, une approche base de données*.
- Pineau, N., Cordelle, S., & Schlich, P. (2003). Temporal Dominance of Sensations: a new technique to record several sensory attributes simultaneously over time. In, *The 5th Pangborn Sensory Science. A Sensory Revolution*. Boston, USA: U.S. Army Natick Solider Center.
- Pineau, N., de Bouille, A. G., Lepage, M., Lenfant, F., Schlich, P., Martin, N., et al. (2012). Temporal dominance of sensations: what is a good attribute list? *Food Quality and Preference*, 26(2), 159-165.
- Pineau, N., & Schlich, P. (2015). 13 - Temporal dominance of sensations (TDS) as a sensory profiling technique. In J. Delarue, J. B. Lawlor & M. Rogeaux, *Rapid Sensory Profiling Techniques*: Woodhead Publishing.
- Porcherot, C., & Schlich, P. (2000). FLASH table and canonical mapping of potato varieties. *Food Quality and Preference*, 11(1/2), 163-165.

Quero, G. M., Fusco, V., Cocconcelli, P. S., Owczarek, L., Borcakli, M., Fontana, C., et al. (2014). Microbiological, physico-chemical, nutritional and sensory characterization of traditional Matsoni: Selection and use of autochthonous multiple strain cultures to extend its shelf-life. *Food Microbiology*, 38(0), 179-191.

Rai, A. K., & Anu Appaiah, K. A. (2014). Application of native yeast from *Garcinia* (*Garcinia xanthochymus*) for the preparation of fermented beverage: Changes in biochemical and antioxidant properties. *Food Bioscience*, 5(0), 101-107.

Rawson, N. E., Gomez, G., Cowart, B. J., Kriete, A., Pribitkin, E., & Restrepo, D. (2012). Age-associated loss of selectivity in human olfactory sensory neurons. *Neurobiology of Aging*, 33(9), 1913-1919.

Rigou, P., Triay, A., & Razungles, A. (2014). Influence of volatile thiols in the development of blackcurrant aroma in red wine. *Food Chemistry*, 142(0), 242-248.

Rizzolo, A., Grassi, M., & Vanoli, M. (2014). 1-Methylcyclopropene application, storage temperature and atmosphere modulate sensory quality changes in shelf-life of 'Abbé Fétel' pears. *Postharvest Biology and Technology*, 92(0), 87-97.

Romano, R., Næs, T., Brockhoff, P. B., Romano, R., Næs, T., & Brockhoff, P. B. (2015). Combining analysis of variance and three-way factor analysis methods for studying additive and multiplicative effects in sensory panel data. *Journal of Chemometrics*, 29(1), 29-37.

Rossini, K., Verdun, S., Cariou, V., Qannari, E. M., & Fogliatto, F. S. (2012). PLS discriminant analysis applied to conventional sensory profiling data. *Food Quality and Preference*, 23(1), 18-24.

Salvador, A., Igual, M., Contreras, C., Martínez-Navarrete, N., & del Mar Camacho, M. Effect of the inclusion of citrus pulp in the diet of goats on cheeses characteristics. *Small Ruminant Research*(0).

Saporta, G., Hatabian, G. (1986). Régions de confiance en analyse factorielle. In E. D. e. al., *Data Analysis and Informatics*: Elsevier Science Publisher B.V.

Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlation. In T. Naes & E. Risvik, *Multivariate analysis of data in sensory science*.

Schlich, P. (1997a). CAP : une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. In A. G. Agro-Industrie, *5e Journées Agro-industries et Methodes Statistiques*. Versailles: 3-5 december 1997.

Schlich, P. (1997b). CAP : une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. . *5èmes journées Agro-industries et Méthodes Statistiques, Versailles, 3-5 Décembre*

- Schlich, P. (2004). L'analyse en variables canoniques des données de profils sensoriels. Applications pour le contrôle de la performance des sujets et pour la comparaison des produits. In, *8e Journées Agro-Industries et Methodes Statistiques*. Rennes (France): ASU Groupe Agro-Industrie.
- Silvano, M. F., Varela, M. S., Palacio, M. A., Ruffinengo, S., & Yamul, D. K. (2014). Physicochemical parameters and sensory properties of honeys from Buenos Aires region. *Food Chemistry*, *152*(0), 500-507.
- Stone, H., & Sidel, J. (2004). *Sensory evaluation practices (Third edition)*. San Diego, California (USA): Elsevier Academic Press.
- Stone, H., Sidel, J. L., Olivier, S., Woolsey, A. L., & Singleton, R. C. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology*, *45*(11), 24-34.
- Sun, S. Y., Gong, H. S., Jiang, X. M., & Zhao, Y. P. (2014). Selected non-Saccharomyces wine yeasts in controlled multistarter fermentations with Saccharomyces cerevisiae on alcoholic fermentation behaviour and wine aroma of cherry wines. *Food Microbiology*, *44*(0), 15-23.

Trząskowska, M., Kołożyn-Krajewska, D., Wójciak, K., & Dolatowski, Z. (2014). Microbiological quality of raw-fermented sausages with Lactobacillus casei LOCK 0900 probiotic strain. *Food Control*, *35*(1), 184-191.

Urdapilleta, I., Ton Nu, C., Saint Denis, C., & Huon de Kermadec, F. (2001). *Traité d'évaluation sensorielle. Aspects cognitifs et métrologiques des perceptions*. Paris (FRA): Dunod.

- W**illiams, A. A., & Arnold, G. M. (1984). A new approach to sensory analysis of foods and beverages. In P. i. F. Research, *4th Weurman Flavour Resaearch Symps*. Amsterdam.
- Wójciak, K. M., Karwowska, M., & Dolatowski, Z. J. (2014). Use of acid whey and mustard seed to replace nitrites during cooked sausage production. *Meat Science*, *96*(2, Part A), 750-756.
- Worch, T., & Punter, P. H. (2015). 14 - Ideal profiling as a sensory profiling technique. In J. Delarue, J. B. Lawlor & M. Rogeaux, *Rapid Sensory Profiling Techniques*: Woodhead Publishing.

Yamamoto, S., Shiga, K., Kodama, Y., Imamura, M., Uchida, R., Obata, A., et al. (2014). Analysis of the correlation between dipeptides and taste differences among soy sauces by using metabolomics-based component profiling. *Journal of Bioscience and Bioengineering*, *118*(1), 56-63.

TABLE DES FIGURES

Figure 1: Mécanismes du goût	6
Figure 2 : Etapes de l'obtention d'une note après stimulation sensorielle	7
Figure 3 : L'ACP des données du Tableau 5.....	22
Figure 4 : Illustration de la projection d'un produit sur un attribut dans un biplot	23
Figure 5 : Biplot des données de l'exemple ($\alpha = 1$)	24
Figure 6 : Distribution des caractéristiques de la SensoBase.....	36
Figure 7 : Histogramme des notes ramenées entre 0 et 10.....	37
Figure 8 : Histogramme du pourcentage de femmes dans la SensoBase	38
Figure 9 : Distribution de l'âge des sujets dans la SensoBase.....	39
Figure 10 : Histogramme du pourcentage de fumeurs dans la SensoBase	41
Figure 11 : Extrait de la Check-list pour la publication d'une méta analyse (QUORUM)	47
Figure 12 : Diagramme de sélection des jeux de données (QUORUM)	47
Figure 13 : Méthodologie adoptée pour les méta-analyses de ce manuscrit	51
Figure 14 : Distribution du pourcentage de zéros par jeu de données et par descripteur	54
Figure 15 : Caractéristiques des 697 jeux de données sélectionnés	56
Figure 16 : Table CAP du jeu de données cheeses	62
Figure 17 : Exemples d'effets de scaling	64
Figure 18 : Dispersion des notes et écarts-types	64
Figure 19 : Effet de scaling et coefficient de régression	65
Figure 20 : Table MAM-CAP du jeu de données cheeses.....	67
Figure 21 : Obtention du coefficient de scaling global pour le sujet i	70
Figure 22 : Synthèse des modèles et des coefficients de scaling.....	75
Figure 23 : Illustration de la différence entre variabilité et discrimination	76
Figure 24 : Différentes étapes de construction d'une carte sensorielle	77
Figure 25 : CVA-biplot associée au jeu de données cheeses	79
Figure 26 : Principes des différentes cartographies des produits.....	80
Figure 27 : Exemple d'ellipse de confiance	82
Figure 28 : Différents types de CVA sur le jeu de données cheeses	84
Figure 29 : Angles entre les descripteurs de l'ACP et la CVA	90
Figure 30: Représentation graphique de la fonction f	99
Figure 31 : Synthèse du travail effectué.....	107

TABLE DES TABLEAUX

Tableau 1 : Réponses possibles de l'animateur de panel à différentes sources d'erreurs d'origine psychologiques.....	9
Tableau 2 : Type d'épreuves sensorielles en fonction de l'objectif de l'étude	10
Tableau 3 : Traitements statistiques en analyse sensorielle	15
Tableau 4 : Caractéristiques des 38 études de profil sensoriel effectuées en 2014	18
Tableau 5 : Moyennes des 6 produits sur 11 descripteurs	22
Tableau 6 : Tableau d'ANOVA d'un modèle à deux facteurs à effet sujet aléatoire	28
Tableau 7 : Tableau d'ANOVA du jeu de données cheeses	28
Tableau 8 : Nombre de répétitions dans la SensoBase.....	37
Tableau 9 : Nombre de participations par sujet	38
Tableau 10 : Répartition des sujets en fonction du genre dans la SensoBase.....	38
Tableau 11 : Répartition de l'âge des sujets dans la SensoBase	40
Tableau 12 : Répartition du niveau d'étude des sujets dans la SensoBase	40
Tableau 13 : Répartition des fumeurs dans la SensoBase	40
Tableau 14 : Nombre de produits par catégorie de produit.....	41
Tableau 15 : Echelles utilisées dans la SensoBase	42
Tableau 16 : Familles de descripteurs dans la SensoBase	42
Tableau 17 : Distribution des fournisseurs de la SensoBase selon le nombre de jeux de données fournis.....	43
Tableau 18 : POURCENTAGES moyens et médians de zéros par modalité sensorielle.....	54
Tableau 19 : Caractéristiques de la base de données obtenue	55
Tableau 20 : F utilisés pour les tests de performances basés sur le MAM	66
Tableau 21 : P-values des tests de Hotelling associées au jeu de données cheeses sur le plan des deux premiers axes.....	78
Tableau 22 : Différents types d'ellipses pour différentes cartes	83
Tableau 23 : Valeurs des critères de comparaison entre ACP et CVA, CVA sans projection de l'effet sujet et MAM CVA	92
Tableau 24 : Synthèse des méta-analyses effectuées.....	102
Tableau 25 : Récapitulatif des analyses associées aux différentes approches.....	118
Tableau 26 : Résumé des recommandations au niveau du protocole de profil sensoriel.....	121
Tableau 27 : Résumé des recommandations au niveau des analyse de données de profil sensoriel	121

ANNEXES

Annexe 1 : Articles.....	133
A. The MAM-CAP Table : a new tool for monitoring panel performances.....	135
B. PCA versus CVA : Comparaison of 422 datasets.....	139
C. Canonical Variate Analysis of Sensory Data.....	147
D. Multiplicative decomposition of the scaling effect in the Mixed Assessor Model.....	171
E. Taking individual differences into account in the use of the scale with MultiMAM.....	187
F. Do we need to replicate in sensory profiling studies.....	207
Annexe 2 : Rapports de méta-analyses.....	221
A. Rapport R1 : ACP contre CVA, méta-analyse des biplots	223
B. Rapport R2 : Comparaison de la MAM-CVA et de l'ACP sur 379 jeux de données de la SensoBase.....	226
C. Rapport R3 : Comparaison de l'ACP et de l'ACP verticale par méta-analyse de la SensoBase	229
D. Rapport R4 : Etude des performances du panel sur 177 jeux de données	234
E. Rapport R5 : Performances des sujets en fonction de leur caractéristiques	239
Performances des panelistes en fonction du genre.....	242
performances des sujets en fonction du statut tabagique	243
performances des sujets en fonction de l'âge	244
F. Rapport R6 : Equivalence entre nombre de sujets supplémentaires et ajout d'une répétition	
245	
Annexe 3 : Aide des packages TimeSens.....	255
A. Aide du package CVA.....	256
B. Aide du package MAMCAP	261

ANNEXE 1 :

ARTICLES



The MAM-CAP table: A new tool for monitoring panel performances



C. Peltier^{a,b,c,*}, P.B. Brockhoff^d, M. Visalli^{a,b,c}, P. Schlich^{a,b,c}

^a CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

^b INRA, UMRI 1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

^c Université de Bourgogne, UMR Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

^d DTU-Informatics, Technical University of Denmark, Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 24 October 2012

Received in revised form 25 June 2013

Accepted 8 July 2013

Available online 24 July 2013

Keywords:

Panel performances

Mixed assessor model

Scaling

ABSTRACT

Assessor performances in sensory analysis are usually represented by three indicators: repeatability, discrimination and agreement. However, assessors can also differ on the range of their scores, the so-called “scaling effect”. Brockhoff, Schlich, and Skovgaard (2013) proposed the mixed assessor model (MAM) which, as the original assessor model (Brockhoff & Skovgaard, 1994), takes this effect into account, but also allows for the product effect to be tested against a new interaction free of the scaling effect.

The present paper proposes a unified system for monitoring assessor and panel performances based on the MAM. In addition to the product effect (tested at panel and individual levels), scaling and repeatability panel heterogeneities are tested and, when significant, assessors who were responsible for this heterogeneity are identified. Further, the pure disagreement is decomposed over assessors allowing for the individuals significantly responsible for it to be highlighted. Those numerous statistics are summarized into a few diagnostics presented in a synthetic single table called the MAM CAP table.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

In descriptive analysis, assessors score products over several attributes in order to describe the sensory characteristics of products. Then, the panel leader needs to monitor individual and panel performances to validate (or not) the results of the experiment.

Discrimination, agreement and repeatability are the three classical performances considered. Several software applications propose statistical tools to perform tests of performances at panel or assessor level. The major issue for the panel leader is to gather essential information from these tests in order to get an overall view leading him to good conclusions. Schlich proposed in 1997 a first version of the Control of Assessor Performances (CAP) table (Schlich, 2007), then revisited it in 2006 for the SensoBase project (<http://www.sensobase.fr>) (Pineau, 2006). This table synthesizes all performances in a unique table, using an intuitive set of colors. Easy to read and to understand, this tool is greatly appreciated by the users. Recently, Neville et al. (2013) extended the CAP table to temporal dominance of sensations (TDS) performances.

In addition to the classical performances, individual scoring effects from scale and ranges are quite common in descriptive analysis. For many reasons, some assessors (even trained ones) can use a greater range of intensities on the scale to rate the products. As a

result, assessors may commit magnitude or crossover errors; some may fail to discriminate the products. Brockhoff et al. (2013) proposed the mixed assessor model (MAM) which takes this effect, the so-called scaling effect, into account:

$$Y_{ijk} = \mu + \alpha_i + \nu_j + \beta_i x_j + d_{ij} + \varepsilon_{ijk}$$

with i : number of assessors, j : number of products, k : number of replicates, α_i : assessor effect, ν_j : product effect, β_i : coefficient of scaling,

$$x_j = y_{j.} - y_{...}, \alpha_i \sim N(0, \sigma_{\text{assessor}}^2), d_{ij} \sim N(0, \sigma_{\text{disag}}^2), \varepsilon_{ijk} \sim N(0, \sigma^2)$$

Stemming from the classical two-way ANOVA model with random subject effect, the MAM decomposes the product*subject interaction into a sum of a “scaling effect” ($\beta_i x_j$) and a “real disagreement” (d_{ij}) (cf. Appendix).

β_i is the coefficient by which we have to multiply the panel deviations to be as near as possible of the panelist deviations. This coefficient of scaling is obtained by linear regression. So, the term $\beta_i x_j$ corresponds to the part of interaction due to a different use of the scale of the assessor, and not to a real disagreement.

The “pure disagreement” d_{ij} is the disagreement freed from the scaling effect. MAM uses this pure disagreement instead of the interaction in the tests of ANOVA, improving products discrimination.

According to the SensoBase study of Brockhoff (2013), a significant scaling heterogeneity was found for 45% of the attributes and 92% of the assessors had at least one significant scaling effect. Thus,

* Corresponding author at: CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France. Tel.: +33 (0) 380681676.

E-mail address: caroline.peltier@dijon.inra.fr (C. Peltier).

almost everybody is concerned by scaling differences. Furthermore, with MAM, the product effect was significant ($\alpha = 0.05$) for 66.1% of the attributes, whereas it was only 60.3% with the classical model and 32.8% are significant for the MAM pure disagreement against 52.4% for the classical interaction. So, the scaling effect is an important phenomenon in descriptive analysis. Taking it into account allows a better description of the assessor scoring behavior.

In this paper, the CAP table is updated, using tests stemming from the MAM, in order to get a fast and synthetic diagnosis tool over all performances. This article begins by a presentation of the new statistical tests, then an example of the new “MAM-CAP” table on a real sensory dataset is exposed, and finally the results are discussed.

2. Materials and methods

The classical individual performances are:

- Discrimination: the ability to detect differences between products,
- Agreement: the fact to agree with the panel on product differences,
- Repeatability: the ability to give the same score to a same product tested multiple times.

These definitions of individual performances can be extended at a panel level based on the ANOVA model used for analyzing the data. In their forthcoming article, Brockhoff et al. (2013) propose a set of performances tests at panel level based on their new MAM model.

Regarding the classical performances, the pure disagreement term (freed from scaling effect) replaced the classical interaction term. It leads to better discrimination and agreement. The panel performance regarding scaling is tested too and gives results about the heterogeneity of scaling (a heterogeneous panel will be significant). Every F-test is conducted with a type 1 error of 0.05.

The Table 1 sums up the statistics of tests for the overall different performances. (Brockhoff et al., 2013).

Furthermore, Brockhoff et al. (2013) indicated that the overall ANOVA table can be explicitly expressed by the individual terms stemming from each assessor. Such an individual decomposition naturally leads to tests for each individual term which are presented in Table 2. Every F-test is conducted with a type 1 error of 0.05.

In Brockhoff (2003) a similar collection of univariate individual performance measures and hypothesis tests were introduced. The strength of the current approach is the explicit expression of the performance measures within the MAM – the MAM was not introduced in Brockhoff (2003). This also means that all the measures and tests given here in the novel MAM-CAP table are based on the simple individual decomposition of the overall ANOVA table, and does not require iterative solutions to multiplicative models as was the case in Brockhoff (2003).

Table 1
Panel performances.

	Tested hypothesis (H_0)	F	Df numerator	Df denominator
Discrimination (F-Prod)	$\forall j, v_j = 0$	$\frac{MS_{product}}{MS_{disag}}$	$J-1$	$(I-1)(J-2)$
Scaling (F-Scal)	$\forall j, \beta_j = 0$	$\frac{MS_{scaling}}{MS_{disag}}$	1	$(I-1)(J-2)$
Disagreement (F-Disag)	$\sigma_{disag}^2 = 0$	$\frac{MS_{disag}}{MS_{error}}$	$(I-1)(J-2)$	$K(J-1)(I-1)$

Table 2

Individual performances. Let's denote by $SS^i_{..}$ the individual contribution for a sum of squares, $v_j^i, \sigma^i_{disag}, \sigma^i_{error}$, respectively product effect, disagreement effect and error term for the assessor i , and $\bar{\sigma}^i_{error}$ the mean of the panel root square of error, the assessor i accepted.

Individual Performance	Tested hypothesis(H_0)	F statistic	Df numerator	Df denominator
Discrimination	$j, v_j^i = 0$	$\frac{SS^i_{product}/(J-1)}{SS^i_{error}/(K-1)}$	$J-1$	$J(K-1)$
Scaling	$\beta_i = 1$	$\frac{SS^i_{scaling}/1}{SS^i_{disag}/(J-2)}$	1	$J-2$
Disagreement	$\sigma^i_{disag} = 0$	$\frac{SS^i_{disag}/(J-2)}{SS^i_{error}/(K-1)}$	$J-2$	$J(K-1)$
Repeatability	$\sigma^i_{error} = \bar{\sigma}^i_{error}$	$\frac{SS^i_{error}/(K-1)}{\frac{1}{I-1} \sum_{i=1}^I SS^i_{error}/(K-1)}$	$J(K-1)$	$(I-1)J(K-1)$

Note that the F-statistic for the individual discrimination is the usual one-way F-statistic for product differences applied to the individual data.

The next paragraph will show an example of MAM-CAP table using a sensory dataset coming from a multi-national project on coffee sponsored by the European Sensory Network (ESN). Focusing on a subset of 16 different coffees profiled by an expert (French) panel of 10 assessors, 13 attributes were scored on a visual analogical scale from 0 to 100. More details on the whole study can be found in European Sensory and Consumer Study: A Case Study on Coffee (European Sensory Network, 1996).

3. Results

The objective of the MAM-CAP table was to present the panel performances as a synthetic table, using the MAM model for testing discrimination, agreement, repeatability and scaling at panel and assessor level.

Each line of the MAM-CAP table corresponds to one attribute. Each cell diagnoses a performance with the help of an easy-to-read “red light – green light” color code: red represents poor performance, green represents good performance. The tests done are those detailed in 2.

The left part of the table presents the panel performances. The first column contains the mean of the panel for each attribute. The three following columns are respectively: F statistics of discrimination (F-Prod), scaling heterogeneity (F-Scal) and disagreement (F-Disa). Then comes the Root Mean Squares of Error (RMSE) which gives indication about repeatability.

Attributes are sorted from the most discriminative to the less discriminative (F-Prod).

The right part of the table concerns individual assessor performances: for each attribute and each assessor, the appropriate cell contains indication about discrimination, agreement, scaling and repeatability.

The discrimination is indicated by the number of dashes in the cell (from zero to three, according to the p-value of the discrimination test: the more dashes, the more discriminative the assessor).

The agreement with the panel is pointed out with the cell color (red if disagreement, green if not). However, if the assessor is not discriminative, the agreement with the panel is not tested, and then the cell color is yellow. A “0” inside a yellow cell occurs when an assessor scores the products (and their replicates) all the same, $\sum_{j=1}^J (y_{ij} - y_{i.})^2 = 0$. In this case, the assessor does not detect differences and can be considered as “not discriminative” for this attribute.

The scaling effect is symbolized by “<” if $\beta_i > 1$ (larger scale use), “>” if $\beta_i < 1$ (smaller scale use) and “|” (non-significant scaling effect).

“? ?” denotes a significant negativity issue (Brockhoff et al., 2013).

Panel Performances						Assessor Performances									
Attribute	Mean	F-Prod	F-Scal	F-Disag	RMSE	S4	S2	S8	S9	S3	S10	S5	S6	S7	S1
EARTH0	8.51	22.79	4.75	0.60	16.99	><	---	<-->	<-->	--	><	lr	--	<-->	--
BITTERT	42.53	20.77	3.21	0.84	15.53	<-->	---	---	---	><	---	><	lr	<-->	--
BURNTF	30.77	20.55	12.16	0.80	16.25	<-->	---	<-->	--	><	<-->	><	><	-- lr	><
BODYMF	38.69	15.48	2.16	0.88	12.32	><	--	--	---	><	--		---	<-->	lr
TOASTO	19.88	7.69	2.86	0.76	20.10	><	><	<>	<>lr	><	<>	--	-- lr	<>lr	--
CHEMO	16.46	6.81	2.21	0.98	16.50	lr		---	<-->		<-->	lr	<-->		><lr
SWEETT	6.31	5.16	3.43	0.77	12.11	--	??	<-->	--	><	0.77	--	lr	<>lr	
FLORALO	11.64	4.99	3.24	0.62	17.41		--	<>	<>lr	><	<>	><	lr		lr
ASTRGNMF	22.41	4.44	2.06	1.42	11.27	---	<>lr	--	--	><	<-->		lr	--	
ACIDT	32.41	4.37	2.41	1.46	13.25	---	---	--	--	--	>---<	<-->lr	lr	><	><
SPICYO	11.70	3.56	3.30	0.79	15.50	--		<>	<>	?--?	<>lr		lr		><
RANCIDF	1.99	2.33	5.10	0.79	6.32	<-->	0	><		><		<-->lr	lr		
TOBACOO	6.70	1.44	3.15	1.00	11.95	??	--			?--?		--	<>lr		
F Rank	-	-	-	-	-	4.29	4.57	4.71	4.96	5	5.14	6	6.14	6.79	7.39

Panel Performances			Assessor Performances	
F-Prod	F-Scal	F-Disa	Product discrimination: --- p<0.01 -- p<0.05 - p<0.1	
P<0.05	P>0.05	P>0.05	Scaling information: : Average scale <> : Significant larger scale >< : Significant smaller scale ? ? : Negative scale	
P>0.05	P<0.05	P<0.05	Agreement information: Yes No No test	
			Repeatability: lr : less repeatable	

Fig. 1. Example of MAM-CAP table and its legend.

Finally, less repeatable assessors are identified with the letters “lr”.

Assessors are sorted for the most discriminative to the least discriminative thanks to the F-Rank: for each attribute, the assessors are sorted by decreasing F-Prod. So, they get a rank per attribute and the mean of these ranks is the F-Rank. So, the smaller the F-Rank, the more discriminative the assessor.

Fig. 1 illustrates results from the MAM-CAP table.

The panel is discriminative for every attribute but TOBACOO. It means that either there is really no difference between products for TOBACOO, or the panel is poorly trained for this attribute. As the panel is in agreement regarding this attribute, the first option seems to be the proper one.

A scaling heterogeneity is observed over all attributes. S1 and S3 seem to use a smaller part of the scale, whereas S8 and S9 a larger one.

Regarding the disagreement, only ASTRGNMF and ACIDT are significant whereas ASTRGNMF, ACIDT, BURNTF and RANCIDF were significant for the interaction with the classical model. This disagreement can be explained by several assessors in disagreement with the rest of the panel (for example, S4, S9 and S10 for ACIDT). However, assessors in disagreement do not imply that the whole panel is in disagreement (TOBACOO for example).

The RMSE gives indication over the panel repeatability: the less repeatable attributes are TOASTO and FLORALO.

Regarding the BODYMF assessor performances, 6 assessors are discriminative (at 5% level) and in agreement with the rest of the panel (green cells). S6 is the only assessor detecting differences between the products in disagreement with the panel (red cell), and S4, S1 and S5 are not discriminative (yellow cells). The heterogeneity of scaling is explained by S7 (resp. S4 and S3) using a larger (resp. smaller) scale than the others.

S2 has put exactly the same note to all products over RANCIDF. A scaling negativity issue appears for SPICYO and TOBACOO in S3 results.

Finally, the order of the assessors on the table (based on F-Rank) shows that the assessor S4 (resp. S1) has been the more (resp. the less) discriminative considering all attributes.

To conclude, the MAM-CAP table shows that this panel is globally well trained. However, the attribute TOBACOO is not discriminant for the products, and should be removed of the study. Then, some individual training could be done in order to correct the negativity issue observed on a few occasions.

4. Discussion

The MAM-CAP table offers an overview of panel performances. However some points should be regarded with special attention.

First, the panel results (agreement, discrimination, scaling) are not directly linked to the individual results. Indeed, the individual tests use only the variability of the scores for one assessor whereas the panel tests use the whole variability. These tests have not the same number of degrees of freedom. Sometimes, it can lead to situations which seem to be contradictory: the panel can be discriminant and in agreement whereas there is no individual discrimination and agreement in the table. For example, the panel is in agreement for FLORALO whereas only S3 is discriminant and in disagreement with the panel. In this case, even if the panel is not really “individually” discriminant and in agreement, it is “collectively” discriminant and in agreement. However, such an attribute would require a retraining of the panelists in the hope of getting more of them discriminative.

Then, assessor repeatability should be considered as relative to the panel repeatability. So, the presence of “lr” in the MAM-CAP table does not mean that an assessor is not repeatable, but that this assessor is less repeatable than the panel. There is no test concerning the panel repeatability. However, it could be compared with the distribution of RMSE obtained by meta-analysis of large number of datasets (Mammasse, 2012).

Finally, the interpretation of negative scaling coefficients represented by “? ?” in the MAM-CAP table can be difficult (real disagreement with the panel, misconception...). Brockhoff et al. (2013) proposes several solutions (removing the observations, using classical model with interaction...). The MAM-CAP table just allows the detection of these negative cases, and let the panel leader decide of the treatment he wants to apply.

Considering the statistical tests, the limit of 0.05 for individual tests can be discussed. As with Bonferroni test, this limit could be divided by the number of subjects, in order to control the error committed by the number of the tests.

Future improvements of the MAM-CAP table will feature multivariate performance indicators.

5. Conclusion

The MAM-CAP table gives visual and synthetic information about performances. It is based on statistical tests stemming from the MAM, a model taking the scaling effect into account, allowing for a better discrimination among products and agreement among subjects.

Consequently, the MAM-CAP table offers the panel leader a fast diagnosis tool about discrimination, agreement, repeatability and scaling of his panel and assessors. This performance diagnosis can be particularly useful during the training phase of the panel.

The MAM-CAP is part of the analyses of TimeSens©. Moreover, the R source code for this table will be freely available on the TimeSens© Website (www.timesens.com).

Appendix

The origin of the mixed assessor model

To understand the origin of the mixed assessor model, let's begin by explaining the classical model of ANOVA.

The usual model of ANOVA is: $Y_{ijk} = \mu + \alpha_i + v_j + c_{ij} + \varepsilon_{ijk}$ with α_i : assessor effect, v_j : product effect, c_{ij} : interaction effect, $\varepsilon_{ijk} \sim N(0, \sigma^2)$.

This model relies on a sum of squares decomposition:

$$SS_{\text{total}} = SS_{\text{prod}} + SS_{\text{subject}} + SS_{\text{inter}} + SS_{\text{error}}$$

The term of interaction, SS_{inter} can be written as:

$$SS_{\text{inter}} = K \sum_{i=1}^I \sum_{j=1}^J ((y_{ij} - y_{i..}) - (y_{j..} - y_{...}))^2$$

with I : number of assessors, J : number of products, K : number of replicates.

Thus, the interaction can be interpreted as the sum of squares of differences between the assessor means deviations and the panel deviations. These differences can be explained by a real disagreement with the panel or a different use of the scale called “scaling effect”.

As an example, let's consider an assessor scoring three products respectively with means 0, 2, 4 and a panel scoring them as 1, 2, 3. Either the assessor can feel a bigger difference between the products than the panel, or agrees with the panel and just uses a larger part of the scale. The mixed assessor model takes this second case into account.

So, if the assessor i differently uses the scale, there exists β_i (scaling coefficient obtained by linear regression of the assessor deviations on the panel deviations) and R_{ij} (pure disagreement) such as:

$$(y_{ij} - y_{i..}) = \beta_i(y_{j..} - y_{...}) + R_{ij}$$

If the assessor uses the scale exactly the same way as the panel, $\beta_i = 1$.

If the scale of the assessor is larger than the panel one, $\beta_i > 1$.

If the scale of the assessor is smaller than the panel one, $0 < \beta_i < 1$.

If $\beta_i < 0$, there is negativity issue and it could be interpreted as a lack of understanding of the scale. This point is discussed in Brockhoff et al. (2013).

Reformulating the term of interaction with this decomposition, the new interaction sum of squares is:

$$SS_{\text{inter}} = SS_{\text{scaling}} + SS_{\text{disag}}$$

$$\text{where } SS_{\text{disag}} = K * \sum_{ij} R_{ij}^2$$

This decomposition leads to the Mixed Assessor Model of Brockhoff:

$$Y_{ijk} = \mu + \alpha_i + v_j + \beta_i x_j + d_{ij} + \varepsilon_{ijk}$$

with α_i : assessor effect, v_j : product effect, β_i : coefficient of scaling,

$$x_j = y_{j..} - y_{...}, d_{ij} \sim N(0, \sigma_{\text{disag}}^2), \varepsilon_{ijk} \sim N(0, \sigma^2)$$

References

- Brockhoff, P. B. (2003). Statistical testing of individual differences in sensory profiling. *Food Quality and Preference*, 14, 425–434.
- Brockhoff, P.B., Schlich, P., & Skovgaard, I.M. (2013) Accounting for scaling differences in sensory profile data: Improved mixed model analysis of variance. *To be submitted*.
- Brockhoff, P. M., & Skovgaard, I. M. (1994). Modelling individual differences between assessors in sensory evaluations. *Food Quality and Preference*, 5, 215–224.
- European Sensory Network (1996). *A European sensory and consumer study: A case study on coffee*. European Sensory Network. Campden and Chorleywood Food Research Association, Chipping Campden, UK.
- Mammasse N., (2012) *Le nombre de sujets dans les panels d'analyse sensorielle : une approche base de données*, Ph.D. thesis, Sciences de l'Alimentation, Université de Bourgogne, France.
- Neville, T., Lepage, M., Rytz, A., Lenfant, F., Martin, N., Schlich, P., & Pineau, N., (2013). Panel performance for temporal dominance of sensations, in preparation.
- Pineau, N. (2006). *La performance en analyse sensorielle. Une approche base de données*. Ph.D. thesis, Sciences de l'Alimentation, Université de Bourgogne, France.
- Schlich, P. (2007). CAP: une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. In 5èmes Journées Agro-industries et Méthodes Statistiques. *Versailles*, 8, 1–10.



Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies

C. Peltier*, M. Visalli, P. Schlich

CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

Université de Bourgogne, UMR Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

ARTICLE INFO

Article history:

Received 7 January 2014

Received in revised form 5 May 2014

Available online xxxx

Keywords:

CVA

PCA

Meta-analysis

SensoBase

ABSTRACT

Although Principal Component Analysis (PCA) of product mean scores is most often used to generate a product map from sensory profiling data, it does not take into account variance of product mean scores due to individual variability. Canonical Variate Analysis (CVA) of the product effect in the two-way (product and subject) multivariate ANOVA model is the natural extension of the classical univariate approach consisting of ANOVAs of every attribute. CVA generates successive components maximizing the ANOVA F-criterion. Thus, CVA is theoretically more adapted than PCA to represent sensory data. However, CVA requires a matrix inversion which can result in computing instability when the sensory attributes are highly correlated.

Based on the analysis of 422 descriptive sensory studies gathered in SensoBase (www.sensobase.fr), this paper compares the maps obtained by covariance PCA and CVA, both performed on significant ($p = 0.1$) attributes for the product effect. Results suggest that 9 times out of 10, PCA and CVA maps are very similar. However, differences between these maps increase with product space complexity. It is thus recommended to analyze and map each sensory modality (texture, aroma...) separately.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

Sensory profiling is the process by which a panel of trained panellists scores the perceived intensities of a number of sensory attributes on a number of products possibly with replications. The resulting data are usually analyzed with Analysis of Variance (ANOVA) to detect differences between products for each attribute taken separately and then with Principal Component Analysis (PCA) to get a product map based on all the attributes simultaneously. Whereas ANOVA takes into account the subject variability around the product means to assess product differences, PCA draws the product map based on product means with no consideration of individual variability. Multivariate Analysis of Variance (MANOVA) is the natural extension of ANOVA to several variables analyzed simultaneously. It allows us to assess statistical significance of product differences in the space generated by the set of attributes being scored. Canonical Variate Analysis (CVA) is a mapping method (cf. Appendix A) based on MANOVA which derives successive components on which the products are as much discriminated as possible. The meaning of discrimination is to separate

product means as much as possible, while individual assessments of a given product are clustered as much as possible close to its product mean. CVA is thus theoretically more adapted than PCA for the multivariate analysis of profiling data. Indeed, Heymann and Noble (1989), Monrozier and Danzart (2001), Noble, Williams, and Langron (1984), Porcherot and Schlich (2000) and Schlich (2004) recommended CVA for the analysis of sensory descriptive data.

Only few comparisons of the two mapping methods have already been reported so far. Martinez and Kak (2001) proved that Linear Discriminant Analysis (a one-way model of CVA) was not always better than PCA to recognize faces. In sensory analysis, Heymann and Noble (1989) showed that PCA and CVA maps provided similar results using wine datasets. Brockhoff (2000) emphasizes the importance of taking into account the variability around the product means with 39 sensory datasets and consequently claimed that CVA was better than PCA. Finally, Monrozier and Danzart (2001) compared PCA, CVA and other different MANOVA tests with resampling and concluded that PCA was very robust whereas complex MANOVA models led to unstable results. These comparisons were carried out with only a few sensory datasets.

This paper presents a comparison of the maps obtained by PCA and CVA on 422 sensory profile datasets. The first objective of this

* Corresponding author. Address: Centre des Sciences du Goût et de l'Alimentation, 9E boulevard Jeanne d'Arc, F-21000 Dijon, France. Tel.: +33 3 80681676.

E-mail address: caroline.peltier@dijon.inra.fr (C. Peltier).

paper is to quantify the differences between PCA and CVA observed on these numerous datasets. For this purpose, criteria for comparing PCA and CVA have been defined. The second objective of this paper is to explain potential differences between PCA and CVA maps in function of the parameters of the datasets (number of products, number of attributes. . .). Our hypothesis is that the more complex the product space, the more different the maps. The complexity of the product space corresponds to its dimensionality which depends on the number of attributes and their level of correlation (the larger the correlations, the smaller the dimensionality). Consequently, decomposing the space into subspaces by sensory modality (texture, smell, taste. . .) is not only intuitively appealing but could also lead to a better agreement between PCA and CVA maps. Such decomposition has been undertaken for datasets showing the largest difference between PCA and CVA maps.

Material and methods

Selection of datasets and attributes

SensoBase (www.sensobase.fr) was created (Pineau, 2006) and maintained by the laboratory of the authors of the present paper. It is a sensory profile database containing 1107 datasets. 422 datasets presenting the following characteristics were extracted from this database:

- Including between 3 and 20 products.
- Including more than 4 subjects.
- Having at least 2 significant ($p = 0.1$) attributes for the product effect.
- Being balanced (same number of evaluations of each product by subject).

These 422 datasets totaled 2130 subjects, 2605 products, 7217 attributes and 1,425,056 observations (scores). For each dataset, covariance PCA and CVA were computed with a R-package developed by the corresponding author. The section 'Comparison with correlation PCA' will address whether a correlation PCA would have given the same results or not. In order to reduce noise in the data, only significant ($p = 0.1$) attributes for the product effect in the two-way ANOVA were included in the analysis.

Number of dimensions used in PCA and CVA

In CVA, a statistical test (Mardia, Kent, & Bibby, 1994) allows us to find the number D of significant dimensions of the product space (cf. Appendix B). Consequently, the first D axes of the CVA theoretically account for all relevant information, the other axes representing only noise. However, many users often prefer to limit their interpretation to the two first axes. Consequently, CVA and PCA plots were compared on the first two dimensions (1–2 comparison), but also in the product space generated by the first D dimensions (1- D comparison).

Map characteristics

CVA or PCA plots are composed of a product representation and an attribute representation. Fig. 1 is an example of CVA plots from a dataset of the SensoBase, having 6 products and 4 significant attributes: Sweet, Salty, Sour and Dry.

In order to be as exhaustive as possible in the comparison, the different elements composing CVA and PCA plots have been listed. They are the same in PCA and CVA plots.

The plot on the left is the representation of the **product configuration**, with points indicating product means. Distances between products represent differences between them. In the example, P_1 appears different from P_5 .

The plot on the right is the representation of attribute configuration, with arrow coordinates being correlation coefficients between attributes and components. The joint interpretation of the two graphs provides a **sensory interpretation**. A product with a large coordinate on the first axis is likely to have large scores on the attributes strongly positively correlated with the first axis. For example, P_5 and P_6 should have larger Salty scores than other products.

The ellipses on the product plot represent the variability of the subject scores around the mean. They delimitate an area where the product means have a 90% probability to be truly located. The overlapping of ellipses gives indications on **pairwise product comparisons**. For example, since P_1 and P_6 ellipses are well separated, they are expected to be different at least on the subspace generated by the first two axes. Note that confidence ellipses can also be built in PCA. Although PCA was computed from the product mean scores, it is possible to project the individual scores as supplementary

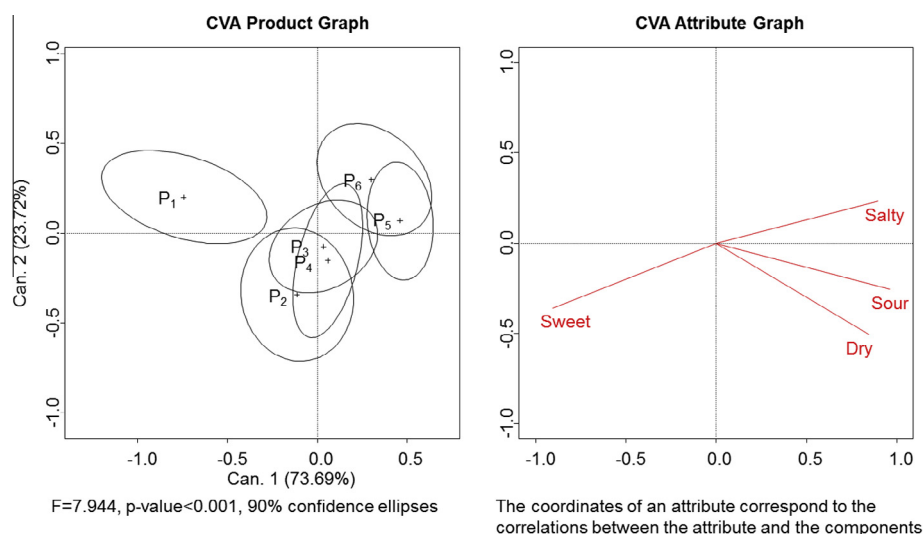


Fig. 1. Example of CVA plots (on the first two components).

elements. Therefore, a covariance matrix related to these projected points can be calculated and used to draw confidence ellipses under a bi-normality assumption (Saporta & Hatabian, 1986).

The MANOVA F is obtained according to the Hotelling–Lawley statistic; other MANOVA statistics exist and are described in Siotani and Wakaki (2006). p -Value of this statistic shows the significance of the multidimensional discrimination test. This F is usually calculated from the entire space, but can also be calculated from fewer components (**Restricted F**) as a natural discrimination indicator of the subspace generated by these components. When the subspace is the map generated by the first two axes, restricted F can be understood as a ratio between product mean distances and average confidence ellipse surfaces.

Criteria for map comparison

The previous section suggests that comparison of PCA and CVA can be based on:

- Similarity of product configurations.
- Similarity of sensory interpretation.
- Output of pairwise product comparisons.
- Value of restricted F .

Any of these properties can be evaluated on both map generated by the two first components (1–2 components) and space generated by the D first components (1– D components). In the latter, product and attribute coordinates have D dimensions and ellipses become ellipsoids. This section presents a methodology for quantifying these elements of comparison

Product configuration

Product configurations can be compared with a RV coefficient. This coefficient is a multidimensional coefficient of correlation, invariant by rotation, dilatation or axial symmetry of the points. If the RV coefficient is close to 1, then the configurations are similar. Contrariwise, an RV coefficient close to 0 indicates no similarity between two configurations. Testing significance of RV coefficient with sensory data was first proposed by Schlich (1996) and discussed by Josse, Pagès, and Husson (2008) who offered a corresponding function within the FactoMineR R-package (Husson, Josse, Le, & Mazet, 2013). The null hypothesis of this test is “RV = 0”. When significant ($p < 0.05$), it means that the observed RV is higher than the RV computed after having permuted the products into one of the two product configurations.

Sensory interpretation

Product positions are interpreted according to attribute directions. To compare PCA and CVA interpretations, the two product configurations were rotated until being superimposed as much as possible with a procrustean rotation (Gower, 1975). Then, the same transformation was applied to the attribute representation to make the attribute directions comparable. For each attribute, the angle (in degrees) between PCA and CVA attribute representations was calculated (Fig. 2).

Two criteria of sensory interpretation were retained for each dataset:

- **The average angle between PCA and CVA attributes** indicating the average interpretation difference.
- **The maximum angle between PCA and CVA attributes** indicating the worst interpretation difference.

Note that an angle of 180 degrees corresponds to the highest possible difference of sensory interpretation. Contrariwise, an angle of 0 corresponds to the smallest possible difference of sensory interpretation.

F discrimination

To quantify the improvement of discrimination from PCA to CVA, the ratio $\frac{F_{CVA}}{F_{PCA}}$ (**F-ratio**) was computed, when F_{CVA} and F_{PCA} are restricted to the space generated by the 2 or D first components. As CVA maximizes discrimination between the products, F_{CVA} should systematically be better than F_{PCA} , and this ratio should be systematically higher than 1.

Pairwise product comparison

Hotelling's T^2 test (Hotelling, 1931) is the multidimensional extension of the usual t -test. It is used to test the multidimensional differences between two products in CVA (or PCA). For each product pair, a Hotelling's T^2 test (restricted to either the first two or the first D components) can be done for both PCA and CVA in order to test if products are significantly ($p < 0.05$) different. A significant p -value of this test generally indicates a separation of the two corresponding confidence ellipses.

If products are declared different (or equal) in both PCA and CVA, then the two approaches agree on that pair. Contrariwise, if one approach found a difference and not the other one, a disagreement is found between them for that pair. To summarize, the comparison between PCA and CVA on one datasets, 3 proportions were computed

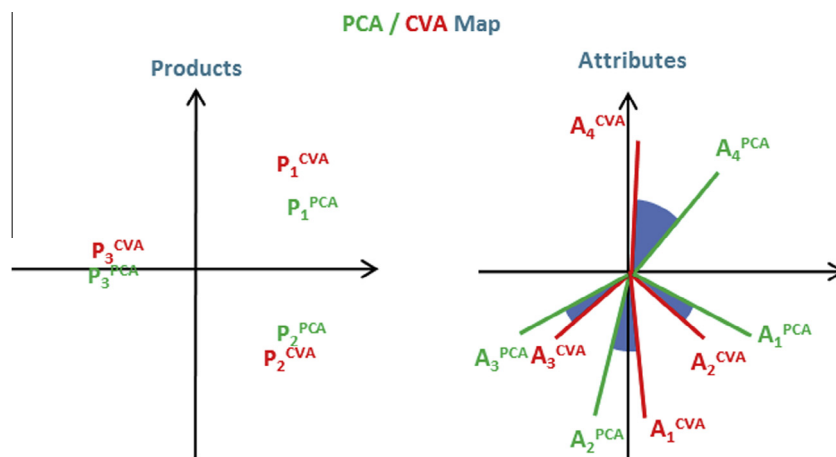


Fig. 2. Angles between PCA and CVA representations of attributes after procrustean rotation.

- The proportion of product pairs with the same diagnosis.
- The proportion of product pairs discriminated in CVA but not in PCA.
- The proportion of product pairs discriminated in PCA but not in CVA.

Finally, the distributions of each of these three proportions were observed.

Investigating origins of differences between maps

Correlations between comparison criteria and datasets parameters

A dataset can be characterized by direct parameters such as the numbers of products, panelists, replicates, attributes and by calculated parameters such as the numbers of significant attributes and significant dimensions. Both direct and calculated parameters could have an influence on differences between PCA and CVA plots.

Correlation coefficients were calculated between these parameters and each comparison criterion. As the studied quantities were not always comparable, the Spearman coefficient of ranks was used. These coefficients were calculated for 1–2 and 1-D comparisons.

The effect of analyzing data by sensory modality

Since one of the paper's hypothesis was that analyzing profiling data by sensory modality (texture, taste, ...) could reduce differences between maps, the 16 datasets with the most different CVA and PCA maps were selected (based on RV criterion only) and new maps by sensory modality were plotted and compared using the different comparison criteria in 1–2 and 1-D space.

The effect of the number of products in differences between maps

The number of products in the dataset could impact the quality of the comparison criteria. Indeed, when using datasets with

a small number of products, it is easy to reach an agreement as measured by RV coefficient. In this case, setting datasets with 3 or 4 products and datasets with more than 10 products on the same study would not be recommended. Consequently, the results were broken down into different ranges of the number of products:

- “Small” for 3 or 4 products.
- “Middle” for 5 to 10 products.
- “Large” for more than 10 products.

The aim of this part is to check that pooling the 422 datasets together does not change any conclusion.

Other questions answered by the comparison methodology

Impact of the collinearity in the data on the differences between maps

CVA involves a matrix inversion. Therefore, if the degree of collinearity in the matrix (correlations among attributes or linear relationships between them) is too high, the inversion can lead to unstable choices of CVA axes; this problem is also known as the “conditioning issue”. The present paper does not test the stability of maps, but looks at the impact of level of collinearity on the differences between PCA and CVA. Indeed, the impact of collinearity on the quality of the CVA map can be an obstacle to the use of CVA. To quantify the degree of collinearity in the data, the condition index (C.I.) was calculated (cf. Appendix C). According to Callaghan and Chen (2008), the degree of collinearity is:

- Weak if the C.I. is lower than 10.
- Moderate to strong if C.I. is between 10 and 30.
- Severe if C.I. is higher than 30.

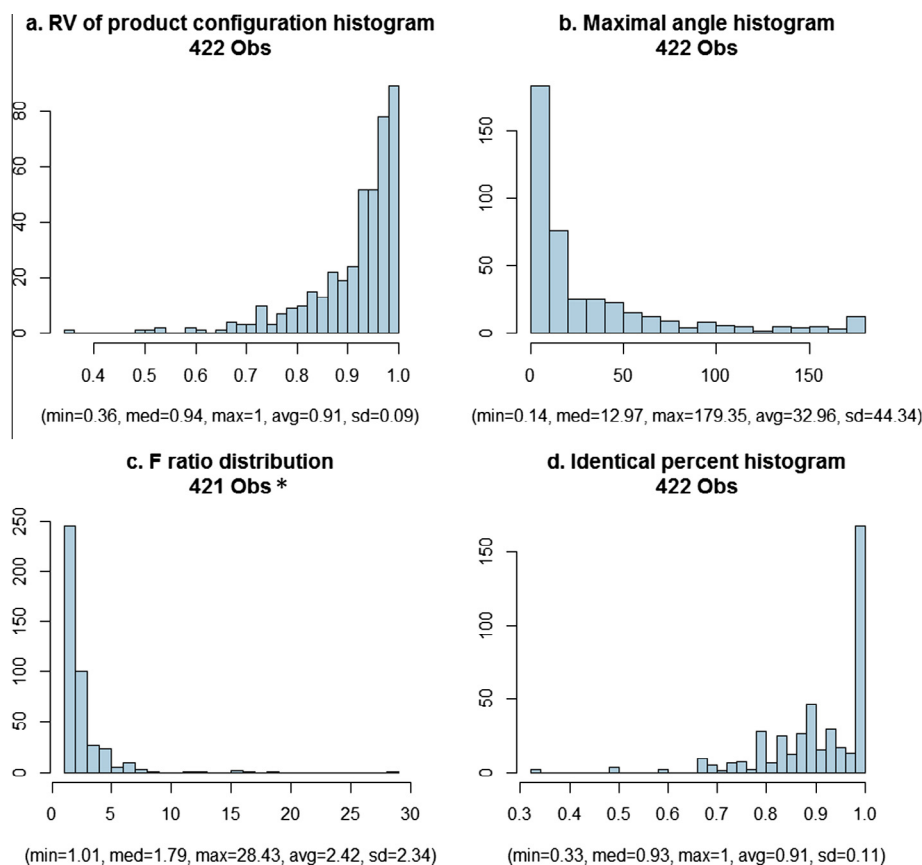


Fig. 3. Distributions of the comparison criteria (on two dimensions).

Consequently, datasets were split into three groups according to their C.I., and the average of each comparison criteria was calculated for each group. Datasets with strong collinearities are expected to have higher differences between two maps.

Comparison with correlation PCA

Correlation PCA gives the same importance to all attributes even if they do not detect differences between products, while covariance PCA weights each attribute to reflect its variance. This is the reason why covariance PCA is generally preferred to correlation PCA for the analysis of product means in sensory profiling. The mapping comparison methodology was applied to compare correlation PCA plots to covariance PCA plots and CVA plots (1–2 and 1-D comparisons).

Results

Quantifying differences between maps

Product configurations

Fig. 3a shows that PCA and CVA product configurations were generally very similar (0.91 in average for 1–2 comparison).

Regarding the 1-D comparison, results were slightly higher (0.61 as minimum, 0.94 in average and 0.95 as median).

In fact, RV coefficient was found significantly different from 0 ($p < 0.05$) in 86.7% of datasets for the 1–2 comparison and 87.2% for 1-D comparison. In other words, about one dataset out of ten had a different product configuration in PCA and CVA.

Sensory interpretation

Fig. 3b and Table 1 show that sensory interpretations are almost identical. Indeed, for the comparison on the first two components, the average angle between an attribute in PCA and in CVA was 8 degrees and maximal angular difference between PCA and CVA was 33 degrees in average. These differences were considerably reduced for the 1-D comparison.

F-ratio

The distribution presented in Fig. 3c shows several very high atypical F-ratios (particularly one of 458). Averaging the data being not pertinent in this case, the use of the median, less sensitive to outliers, was preferred. Considering this median, F was 1.8 times better in CVA compared to PCA. The F-ratio in the 1-D comparisons follows almost the same distribution (1.64 as median and 458 as maximum). The restricted F was about two times higher in CVA compared to PCA.

Product pairwise interpretation

Fig. 3d indicates the distribution of the percentages of product pairs with the same PCA–CVA diagnosis. 167 datasets have the same diagnoses in PCA and CVA for all the product pairs. For 1–2

Table 1

Distribution of the average and maximum angles between PCA and CVA attributes.

	Min	1st Q	Med	3rd Q	Max	Avg	Sd
1–2 Average angle	0.14	2.24	4.39	9.46	73.00	8.04	10.05
1-D Average angle	0.14	2.08	3.24	4.79	25.83	3.84	3.00
1–2 Maximum angle	0.14	4.70	12.97	43.62	179.40	32.96	44.34
1-D Maximum angle	0.14	2.89	5.36	8.63	62.08	7.73	8.25

(and 1-D) comparison, the averages of the different proportions presented in section ‘Pairwise product comparison’ were:

- 91.04% (and 93.59%) for the proportions of product pairs with the same diagnosis.
- 3.63% (and 2.60%) for the proportions of product pairs discriminated in CVA but not in PCA.
- 5.33% (and 3.81%) for the proportion of the product pair discriminated only in PCA.

These results show that generally, PCA and CVA give similar diagnosis for product pairwise interpretation. Note however that, surprisingly, CVA does not seem to discriminate more often product pairs than PCA. A possible interpretation of these results will be presented in the ‘Discussion’ section.

Interpretation and reduction of the differences between CVA and PCA

Relations between datasets parameters and comparison criterions

Table 2 suggests that, the more there are products, significant attributes and significant dimensions, the more the maps are different. Furthermore, the more subjects and more replicates, the more identical the maps seem to be.

Surprisingly, the presence of collinearity did not impact the differences between maps as much as we had expected. It seemed to be less important than the number of significant dimensions. F_{CVA}/F_{PCA} was the only criterion strongly correlated to the conditioning index (0.80).

1–2 Comparison results were generally the same for the 1-D comparisons. The strongest correlation is still between the F-ratio and the condition index (0.79). However, the percentage of product pairs with the same PCA–CVA diagnosis is now only correlated with the number of products and the number of subjects (–0.35 and 0.21). Furthermore, the correlations regarding D and the comparison criteria were less strong than previously (–0.38 for the RV, 0.33 for both angles, 0.18 for F-ratio and NS for the percent of product pairs with same PCA–CVA diagnosis). The reason could be that when all significant dimensions of product space (D) are taken into account in the comparison, the number of these dimensions does not impact differences between maps as much as in 1–2 comparison. It shows the importance of considering the components 3 to D.

Table 2

Spearman correlation coefficients between dataset parameters and comparison criteria (1–2 comparison).

	Product RV	Average angle between PCA and CVA attributes	Maximum angle between PCA and CVA attributes	F_{CVA}/F_{PCA}	Percent of product pairs with same PCA–CVA diagnosis
Number of products	–0.32	0.44	0.56	–0.14	–0.42
Number of subjects	0.12	–0.12	NS	–0.14	0.14
Number of attributes	–0.28	0.29	0.33	0.52	–0.18
Number of replicates	0.20	–0.21	–0.23	NS	0.12
Number of significant attributes	–0.40	0.43	0.57	0.67	–0.27
Condition index (C.I.)	–0.32	0.28	0.32	0.80	–0.14
Significant dimension of product space D	–0.50	0.55	0.70	0.36	–0.38

NS indicates a non-significant ($p > 0.05$) coefficient of correlation ($n = 422$).

Table 3
Comparison criteria for complete analysis and analysis by sensory modality.

	Average of comparison criteria for the complete analysis (16 maps)		Average of comparison criteria for the analysis by sensory modality (55 maps)	
	1–2 comparison	1-D comparison	1–2 comparison	1-D comparison
Product RV	0.80	0.85	0.95	0.96
Maximum angle between PCA and CVA attributes	41.50	9.43	13.01	6.82
Average angle between PCA and CVA attributes	13.65	6.46	5.16	3.24
Percent of product pairs with same PCA–CVA diagnosis	90.24%	92.95%	92.98%	94.34%

Table 4
Comparison criteria related to the number of products of the datasets.

Number of products	1–2 Components				1–D components			
	Small	Medium	Large	Total	Small	Medium	Large	Total
Number of datasets	136	227	59	422	136	227	59	422
Dimension of product space	2.1	3.77	4.71	3.36	2.1	3.77	4.71	3.36
Product RV	0.94	0.90	0.89	0.91	0.95	0.93	0.93	0.94
Maximum angle between PCA and CVA attributes	11.7	42.5	45.5	33.0	4.1	7.9	15.5	7.7
Average angle between PCA and CVA attributes	4.2	9.6	10.9	8.0	2.9	3.9	6.9	3.8
Percent of product pairs with same PCA–CVA diagnosis	93.9	90.1	88.2	91.0	93.4	94.1	91.9	93.6
Median of F_{CVA}/F_{PCA}	1.8	1.9	1.4	1.8	1.75	1.73	1.27	1.6

Finally, a high complexity of the product space (represented by a large number of significant attributes and significant dimensions of product space) and a large number of products could be related to differences between PCA and CVA plots in both 1–2 and 1-D comparison. The strong correlation between F-ratio and condition index will be developed in sections ‘Impact of collinearity on CVA maps’ and ‘Discussion’.

Reducing differences between maps in CVA and PCA

The 16 datasets whose RV coefficients for product configurations in CVA and PCA were not significantly different from 0 and presenting enough attributes in several sensory modalities were selected. The comparisons of the 55 resulting CVA and PCA maps showed that 50 out of them became not significantly different by the RV criterion. This result was validated by every criterion (Table 3).

Note also that the analysis by sensory modality in 2 dimensions seems to give more similarity between CVA and PCA than complete analysis on D dimensions (particularly for similarity of product configurations).

Plotting maps by sensory modalities definitely seems to give more similarities between PCA and CVA maps.

The effect of the number of products on differences between maps

Table 4 shows that the product number does impact all comparison criteria: the more there are products, the more different the maps are, as concluded in section ‘Relations between datasets parameters and comparison criterions’.

This table also confirms the conclusions of section ‘Quantifying differences between maps’, even for datasets with a large number of products. Consequently, pooling together the 422 datasets did not led to misinterpretations.

Other questions answered with the same methodology

Impact of collinearity on CVA maps

78.2% of the matrices had a weak degree of collinearity and 2.8% had a “severe” degree of collinearity. The distribution of condition index is given in Table 5.

Apart from the F-ratio, Table 6 shows that differences between maps increased slightly with the conditioning issue.

Table 5
Distribution of condition index.

Min	1st Dec.	1st Qu.	Med	3rd Qu.	9th Dec.	Max	Avg
1.36	2.85	4.21	6.59	9.52	14.8	1167	11.76

The less acute the conditioning issue is, the more similar the product configuration and the product pairs with same PCA–CVA diagnosis in 1–2 and 1-D comparison are. The impact of conditioning issue on angles for sensory interpretation is less clear. In the 1–2 comparison, the variations of angles seem to be linked with the number of dimensions of the product space. This hypothesis is validated by the strong correlation (0.55 and 0.7 in Table 2) between these angles and D. In 1-D comparison, the average angle seems to be higher when there is conditioning issue.

The F-ratio was equal to 1.6 on average for the datasets with a weak degree of collinearity, while it was 6.29 for the datasets with severe collinearity (1–2 comparison). This is in accordance with the strong correlation between the condition index and discrimination ratio. Furthermore, the outlier of F-ratio (458) corresponds to the dataset with the outlying condition index of 1167. These results will be discussed in ‘Discussion’ section.

Note that several regularization procedures could be proposed to cope with possible quasi-collinearities problem in CVA, for example PLS discriminant analysis (Rossini, Verdun, Cariou, Qannari, & Fogliatto, 2012).

Correlation PCA

The same calculations were done to compare correlation PCA, covariance PCA and CVA (Table 7).

Firstly, covariance PCA is more similar to CVA than correlation PCA, which may be seen as an additional justification of using covariance rather than correlation PCA. This result was expected since correlation PCA, unlike CVA and covariance PCA, starts by setting all the attributes to the same variance. Secondly, it seems that there are as much differences between covariance PCA and CVA as between covariance PCA and correlation PCA. Theoretically, CVA is better suited to sensory profiles than covariance PCA, which is better than a correlation PCA. Practically, these maps are generally very similar.

Table 6

Comparison criteria related to the condition index (C.I.) of the datasets.

Conditioning issue	1–2 Components			1-D components		
	Weak	Moderate	Severe	Weak	Moderate	Severe
Number of datasets	363	80	12	363	80	12
Dimension of product space	3.3	3.64	3.08	3.3	3.64	3.08
Product RV	0.92	0.89	0.88	0.94	0.92	0.89
Maximum angle between PCA and CVA attributes	30.9	44.7	16.8	8.1	6.19	7.7
Average angle between PCA and CVA attributes	7.9	9.0	6.4	3.7	3.9	5.5
Percent of product pairs with same PCA–CVA diagnosis	91.2%	90.8%	86.1%	93.8%	94.6%	90.0%
Median of F_{CVA}/F_{PCA}	1.6	3.14	6.29	1.47	2.83	5.91

Table 7

Results of comparison of covariance PCA (PCA), correlation PCA (PCA cor) and CVA.

	PCA vs. CVA		CVA vs. PCA cor		PCA vs. PCA cor	
	1–2	1-D	1–2	1-D	1–2	1-D
Product RV	0.91	0.94	0.86	0.89	0.92	0.94
Maximum angle between PCA and CVA attributes	32.96	7.73	38.36	8.77	28.79	6.99
Average angle between PCA and CVA attributes	5.9	3.84	9.68	4.30	7.10	3.31
Percent of product pairs with same PCA–CVA diagnosis	91%	93.59%	89.35%	92.48%	93%	96.15%

Discussion

In this paper, a methodology to compare PCA and CVA representations was defined. Its originality is to take into account all elements of the representations (product configuration and sensory interpretation, but also discrimination and product pair differences).

Results show that PCA and CVA plots are often (9 time out of 10) very similar whatever the number of components (2 or D) considered. However, one may notice that the differences are still reduced when D components are considered instead of only two.

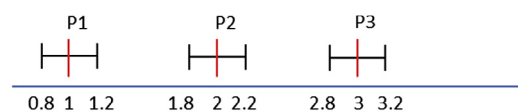
A surprising result was that F-ratio appears to be strongly correlated to the conditioning. The worst the condition index, the better the F-ratio. This result means that the more correlated the attributes, the more CVA can discriminate products. Besides, it is well known that strong collinearities theoretically result in map instability; however, we did not observe larger differences in product maps between CVA and PCA with collinearities. Finally, a large condition index can artificially increase the MANOVA F but does not critically impact the map.

Another unexpected result was that CVA appears not to be better than PCA in pairwise product differences. It can be explained by the fact that the maximization of the average discrimination in CVA does not systematically imply a larger number of differences between products. In other words, to increase the general discrimination, CVA can find a map where one product would be very discriminated from the others, but the other products might be not discriminated from each other. This is an extension of the one-dimensional case, where a large F does not necessarily imply a better discrimination between products taken 2 by 2 (cf. Fig. 4).

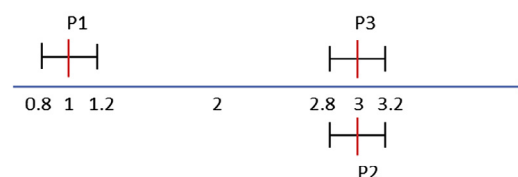
An alternative to the classical representation with a product graph and a distinct attribute graph exists: the biplot. It consists in representing the products and the attributes on the same graph. With this method, projections of products on attribute arrows approximate the scores obtained. Biplots exist in PCA but also in CVA (Gower, Lubbe, & Le Roux, 2011). Consequently, further investigations could extend the comparison of PCA and CVA to biplots. The conclusions of such investigations are expected to be the same as the present paper.

Finally, it is difficult to answer the question “Is CVA better than PCA?” As the underlying “truth” of data is unknown, defining the best product configurations or the best sensory interpretation is impossible *a priori*. On the one hand, PCA is based on the product

Case 1: $F = 12.5$ and 3/3 different pairs of products



Case 2: $F = 16.67$ and 2/3 different pairs of products



Red marks represent product averages. Each product is scored by two subjects (black marks).

Fig. 4. Illustration of the fact that higher F does not imply better differences between products.

average table and could be more stable than CVA. On the other hand, CVA has the advantage to use a statistical model in line with the ANOVA approach. Consequently, CVA can be used to map the levels of any factor in any MANOVA model and by specifying the proper error term. Furthermore, the conditioning issue seems not to have a strong impact on the plots.

When PCA and CVA give the same map, this map could be recognized as a good map to describe products, as it represents both product variability and discrimination. This work showed that reducing dimensionality of the data improved similarities between CVA and PCA. A practical and sensory meaningful way of reducing dimensionality is to analyze the data by sensory modality. This practice seemed to discard most difference between PCA and CVA and besides offers maps much easier to interpret by the sensory analyst. If maps by modality are similar, it becomes legitimate to draw a single map by pooling these sensory modalities; otherwise, it is much better to draw interpretation thanks to the maps by modality. In the latter case, quite often a two-dimensional map for each sensory modality may be sufficient, whereas a joint analysis may require more than 2 axes to recover the variation of the data. This entails the examination of several maps, which is not an easy task. If the user needs to determine the links between the attributes of different sensory modalities, dedicated three-way

multivariate techniques should be used for highlighting the product differences either common to the sensory modalities or specific to each of them.

Conclusion

PCA of the product mean scores is the usual technique for mapping products in quantitative descriptive analysis, but does not take into account the variability around the means. CVA is another method of mapping, based on MANOVA, which maximizes the discrimination instead of variability. This study has not proved that CVA is better than PCA (or inversely). The results show that using PCA extensively in sensory analysis does not result in misinterpretation and loss of information: CVA and PCA are often equivalent whatever the number of components (2 or D) considered.

However some datasets present some differences between their PCA and CVA representations. These differences can be explained by a high number of products and a strong complexity of the product space (number of significant attributes and number of significant dimensions). Consequently, to avoid different conclusions between PCA and CVA representations, plotting maps by sensory modality (texture, odor, taste...) is recommended.

Appendix A. Canonical Variate Analysis

If a dataset contains I products, J subjects, K replicates and D attributes, let us denote PROD as the I^*D matrix of product effects, and INTER the IJ^*D matrix of interaction effects:

$$\text{PROD} = \begin{bmatrix} (\bar{y}_1^1 - \bar{y}^1) & \dots & (\bar{y}_I^D - \bar{y}^D) \\ \vdots & \ddots & \vdots \\ (\bar{y}_I^1 - \bar{y}^1) & \dots & (\bar{y}_I^D - \bar{y}^D) \end{bmatrix} = (\bar{y}_i^d - \bar{y}^d)_{i=1..I, d=1..D}$$

$$\text{INTER} = (\bar{y}_{ij}^d - \bar{y}_i^d - \bar{y}_j^d + \bar{y}^d)_{i=1..I, j=1..J, d=1..D}$$

where \bar{y}^d is the average of the scores for attribute d ; \bar{y}_i^d the score average for attribute d and product i ; \bar{y}_j^d the score average for attribute d and subject j ; and \bar{y}_{ij}^d the score average for attribute d , subject j and product i .

Then, let B and W be the D^*D matrices of covariance of the product and interaction effect:

$$B = \frac{1}{(I-1)} (\text{PROD})' (\text{PROD})$$

$$W = \frac{1}{(I-1)(J-1)} (\text{INTER})' (\text{INTER})$$

If W is invertible, then $W^{-1}B$ can be described as a weighting of the differences between products by the inverse of the disagreement between the subjects (as a parallel, $W^{-1}B$ should be considered as a multidimensional extension of the calculation of the F-statistic in univariate ANOVA which is the ratio between the between variance and within variance). Multidimensional information is extracted from the matrix by computation of the eigenvalues ($\lambda_1, \dots, \lambda_D$) and eigenvectors $v = (v_1, \dots, v_D)$ so called **canonical variates**. These canonical variates maximize the discrimination of the products. To plot the CVA product graph, the table of product means is projected on the canonical variables. Then, the projection of product scores on the first two canonical variates ($\text{PROD}^*(v_1 v_2)$) gives the coordinates of the products in the 1–2 plot. For each product, individual observations are also projected on the canonical variables. With the obtained scatterplot, confidence ellipses can be plotted around them.

Attributes coordinates are the between product correlation coefficients between attributes scores and canonical variates.

Appendix B. Test of dimensionality in CVA

The dimension of the product space is obtained thanks to a test based on the eigenvalues of $W^{-1}B$. This test, presented in [Mardia et al. \(1994\)](#), consists in checking step by step if the last axis (then the two last axes and so on) can be removed from the data without significant loss of information. Indeed, if the last r dimensions are noise in the data, the sum of squares of the r last eigenvalues can be approximated by a χ^2 distribution and a p -value can be obtained.

Appendix C. Calculation of the condition index

Conditioning measures the impact of perturbations of an initial matrix on the inverse of this matrix and corresponds to the degree of collinearity in this matrix. Let M be a matrix and E a matrix of little perturbations. If M has a weak degree of collinearity, $(M+E)^{-1}$ will be very close to M^{-1} . If M has a severe degree of collinearity, $(M+E)^{-1}$ could be very different from M^{-1} and there would be a strong conditioning issue. The conditioning measures the degrees of collinearity in the data making the inversion unstable. When M is symmetric definite positive, conditioning is measured by $\sqrt{\lambda_1/\lambda_{\text{Natt}}}$ (the condition index) where $\lambda_1 > \dots > \lambda_{\text{Natt}}$ are the eigenvalues of M .

In CVA, the M matrix is W .

References

- Brockhoff, P. (2000). Multivariate analysis of sensory profile data: Is CVA better than PCA? *Levnedsmiddelkongres*, Copenhagen, Denmark.
- Callaghan, K., & Chen, J. (2008). Revisiting the collinear data problem: An assessment of estimator 'ill-conditioning' in linear regression. *Practical Assessment, Research and Evaluation*, 13(5).
- Gower, J. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1).
- Gower, J., Lubbe, S., & Le Roux, N. (2011). *Understanding biplots*. John Wiley and Sons, Ltd..
- Heymann, H., & Noble, A. C. (1989). Comparison of canonical variate and principal component analyses of wine descriptive analysis data. *Journal of Food Science*, 54(5), 1355–1358.
- Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics*, 2, 360–378.
- Husson, F., Josse, J., Le, S., & Mazet, J. (2013). *FactoMineR, multivariate exploratory data analysis and data mining with R. R package version 1.25*. <<http://CRAN.R-project.org/package=FactoMineR>>.
- Josse, J., Pagès, J., & Husson, F. (2008). Testing the significance of the RV coefficient. *Computational Statistics and Data Analysis*, 53, 82–91.
- Mardia, K. V., Kent, T. J., & Bibby, J. M. (1994). Multivariate analysis (pp. 339–344).
- Martinez, A., & Kak, Avinash C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2).
- Monrozier, R., & Danzart, M. (2001). A quality measurement for sensory profile analysis. The contribution of extended cross-validation and resampling techniques. *Food Quality and Preference*, 12(5–7), 393–406.
- Noble, A. C., Williams, A. A., & Langron, S. P. (1984). Descriptive analysis and quality ratings of 1976 wines from four Bordeaux communes. *Journal of the Science of Food and Agriculture*, 35(1), 88–98.
- Pineau, N. (2006). La performance en analyse sensorielle, une approche base de données. Thesis.
- Porcherot, C., & Schlich, P. (2000). FLASH table and canonical mapping of potato varieties. *Food Quality and Preference*, 11(1–2), 163–165.
- Rossini, K., Verdun, S., Cariou, V., Qannari, E. M., & Fogliatto, F. S. (2012). PLS discriminant analysis applied to conventional sensory profiling data. *Food Quality and Preference*, 23(1), 18–24.
- Saporta, G., & Hatabian, G. (1986). Régions de confiance en analyse factorielle. In E. D. et al. (Eds.), *Data analysis and informatics*. Elsevier Science Publisher B.V.
- Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. In T. Naes & E. Risvik (Eds.), *Multivariate analysis of data in sensory science* (pp. 259–306). Elsevier Science B.V.
- Schlich, P. (2004). L'analyse en variables canoniques des données de profils sensoriels. Applications pour le contrôle des performances des sujets et pour la comparaison des produits. 8èmes journées Européennes Agro-Industrie et Méthodes Statistiques (AGROSTATS), Agrocampus, Rennes.
- Siotani, M., & Wakaki, H. (2006). Contributions to multivariate analysis by Professor Yasunori Fujikoshi. *Journal of Multivariate Analysis*, 97, 1914–1926.

1
2
3
4
5
6
7
8
9
10
11

Canonical Variate Analysis of sensory profiling data

C. Peltier^{1,2,3,4}, M. Visalli^{1,2,3}, P. Schlich^{1,2,3}

¹CNRS, Centre des Sciences du Goût et de l'Alimentation, Dijon, France

²INRA, CSGA (UMR 1324), France ;

³Université de Bourgogne, France

⁴E2S, Dijon, France

Corresponding author:

Caroline Peltier

caroline.peltier@dijon.inra.fr

12 **Abstract**

13 Principal Component Analysis (PCA) of product mean scores is generally used to obtain a
14 product map from sensory profiling data. However, this approach does not take into account
15 the variance of the product mean scores due to the individual panelist variability.

16 Therefore, Canonical Variate Analysis (CVA) of the product effect in the two-way
17 multivariate ANOVA (MANOVA) should be considered as a natural alternative analysis to
18 PCA. Indeed, it is the extension of the classical univariate approach used for the analysis of
19 each descriptor separately. This analysis generates successive components maximizing
20 product discrimination as measured by the usual Fisher statistics in ANOVA. The outcome of
21 CVA is a product map maximizing product separation, while gathering individual evaluations
22 of the same product as close as possible.

23 This paper is an introduction to the rationale of CVA applied to sensory profiling data. In
24 addition, it offers an R package (CVApack) and instructions for running it.

25 **Practical applications**

26 This paper introduces a mapping method which allows visualizing graphically the results of a
27 two way MANOVA. This method can be used in sensory analysis in order to describe product
28 discrimination. Theoretically more adapted to sensory data than PCA, CVA can also be used
29 in other fields of application with a few adaptations.

30 **Keywords**

31 mapping, sensory analysis, CVA, PCA

32 1. Introduction

33 In sensory profiling, trained subjects taste products and score the perceived intensities of a set
34 of sensory descriptors in order to find out which descriptors generate differences among
35 products. In practice, Analysis Of Variance (ANOVA) is used to establish significance of
36 product differences. The principle of ANOVA is to assess whether the ratio between product
37 differences and subject variability is large enough to declare the existence of significant
38 differences among products. After having analyzed each descriptor independently by
39 ANOVA, most sensory analysts perform a Principal Component Analysis (PCA) of the
40 product mean scores to obtain a bi-dimensional product map delivering a visual representation
41 of the product differences generated by the whole set of descriptors.

42
43 However, PCA being done on the product mean scores, no consideration is given to the
44 variance of these means, which express both subject heterogeneity (disagreement between
45 them) and individual un-repeatability. For a single attribute, this practice studies the mean
46 scores of the products without considering their variances, i.e. forgetting to run an ANOVA
47 on those data.

48 Canonical Variate Analysis (CVA) is a multivariate technique related to Multivariate Analysis
49 of Variance (MANOVA) which allows overcoming PCA limitation (Mardia et al. 1989).
50 Noble et al. seems to be the first to recommend its use with sensory profiling data (Noble
51 1984). Later, Porcherot and Schlich described in more details the theoretical advantages of
52 CVA over PCA on sensory data (Porcherot and Schlich 2000). Although it was recently
53 shown (Peltier et al. 2014) that PCA and CVA did not differ drastically in their practical
54 outcomes, CVA is theoretically more adapted to sensory data. Until now, CVA has not been
55 widely used by the sensory community. We speculate that this is mainly due to the lack of
56 available software supporting it.

57
58 This paper aims to be a tutorial for using CVA on profiling data. The first part of this paper
59 describes the structure of a profile sensory dataset. The second part presents the concept of
60 discrimination and how it can be transposed to multivariate analysis. The third part explains
61 step by step how to obtain a CVA map. The final part highlights CVA specificities. R-
62 package can be downloaded from <http://www.timesens.com> and R-code is available in
63 Appendix 1.

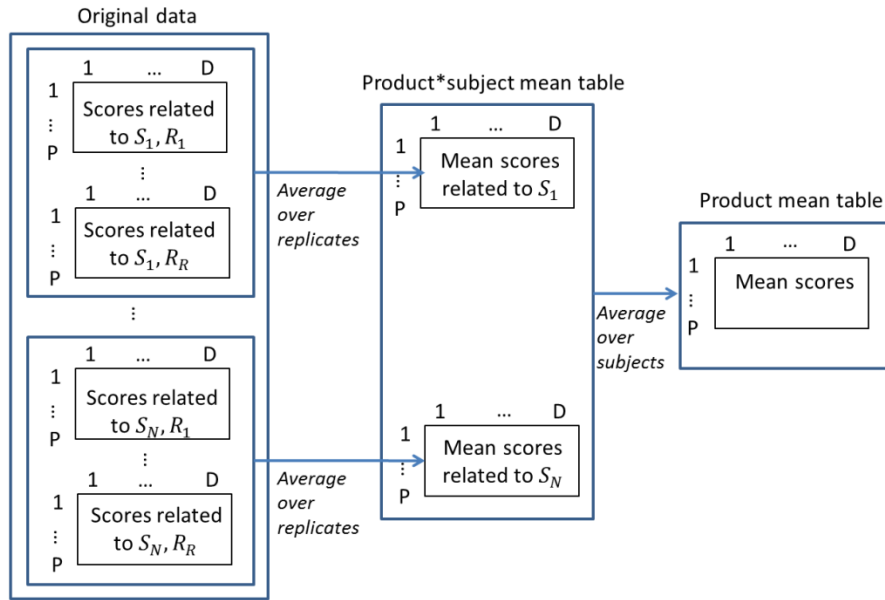
64 2. Sensory profiling data

65 2.1. Structure of a classical dataset

66 Let N subjects score P products on D descriptors R times (replicates). The dataset can be
67 presented in a table where each line (observation) contains the scores given by one subject to
68 one product on one replicate, and each column (variable) contains the scores given to one
69 descriptor. As presented in Figure 1, the table of product by subject means (resp. product
70 means) is deduced from the original data by averaging over the replicates (resp. over the
71 replicates and the subjects).

72 FIG 1. STRUCTURE OF PROFILE DATA: P products (P_1, \dots, P_P) profiled by N subjects
 73 (S_1, \dots, S_N) on D descriptors (D_1, \dots, D_D) in R replicates (R_1, \dots, R_R)

74



75 Mathematically, we can denote then by y_{pnr}^d the score for the descriptor $d = 1..D$, given by
 76 the n^{th} ($n = 1..N$) panelist when he tasted the product p ($p = 1..P$) at the r^{th} time ($r =$
 77 $1..R$). The original data is then a matrix with dimensions $NPR \times D$. The product subject
 78 mean table is a matrix with dimension $NP \times D$ where each coefficient is defined as follows:
 79

$$\bar{y}_{pn}^d = \frac{1}{R} \sum_{r=1}^R y_{pnr}^d$$

80 Finally, the dataset product x descriptor is a $P \times D$ matrix with coefficient defined as follows:

$$\bar{y}_p^d = \frac{1}{N} \sum_{n=1}^N \bar{y}_{pn}^d$$

81

82 2.2. Missing values

83 Missing values in sensory profiling are most often caused by a few subjects having missed
 84 one or several sessions. Several solutions can be proposed to handle missing values,
 85 depending on their nature. When the dataset is still complete despite the presence of missing
 86 values (all subjects have scored all products at least once); no missing values appear in the
 87 product*subject table. Therefore, the same weight is given to a subject with and without

88 missing values. When the presence of missing values makes the dataset incomplete and
 89 affects the product*subject table, these missing values can be imputed (by the panel average
 90 or a combination of panelist and panel average in accordance with the chosen statistical
 91 model). More advanced techniques exist but are not discussed in this paper. Some examples
 92 of statistical methods handling missing values are available in Husson and Josse (2013) or
 93 Grung and Manne (1998).

94 **2.3. Dataset used as example**

95 In the following, a sample dataset will illustrate the developed concepts. This dataset is
 96 extracted from the SensoBase, a sensory database (www.sensobase.fr) containing more than 1
 97 000 profiling datasets. Six products (cheeses) were scored on a continuous scale (from 0 to
 98 10) on 12 descriptors by 16 subjects in triplicate. The dataset is balanced and can be found in
 99 the CVA R-package.

100 **3. Conceptual explanation of discrimination**

101 **3.1. Concept of unidimensional discrimination**

102 **Discrimination** is the ability to detect sensory differences between products at a panel level.
 103 Discrimination differs from product variability because it takes the individual variability into
 104 account.

105 Figure 2 illustrates the difference between the two concepts: the two situations have equal
 106 product mean scores (product variability) but very different distribution of individual scores
 107 around these means (product discrimination).

108 **FIG. 2. TWO SITUATIONS WITH THE SAME PRODUCT VARIABILITY BUT**
 109 **DIFFERENT SUBJECT VARIABILITY**

Two situations with same product variability but different subject variabilities

Situation 1: Strong subject variability suggesting that P1 and P2 are not different



Situation 2: Weak subject variability suggesting that P1 and P2 are different



Two products have been scored by 6 subject in both situations.

Small lines represent the subject scores whereas large lines represent the average.

110 Discrimination can be defined mathematically in the framework of the ANOVA model used
 111 to analyze the data. In sensory profiling, it is fairly well accepted to use a two-way mixed
 112 model with product being a fixed effect, whereas subject is a random one (Næs and Langsrud
 113

114 1998). In this model, ANOVA theory indicates that discrimination consists in comparing the
 115 product mean square ($MS_{Product}$), which is proportional to the variance of the product means
 116 scores, to the product by subject interaction mean square ($MS_{Product*Subject}$), which is
 117 expressing the disagreement between subjects, using the Fisher statistic:

$$118 \quad F_{Product} = \frac{MS_{Product}}{MS_{Product*Subject}}.$$

119 For a fixed descriptor d, the mathematical expressions of $MS_{Product}$ and $MS_{Product*Subject}$ are
 120 as following:

$$MS_{Product}^d = \frac{N \cdot R}{P - 1} \sum_{p=1}^P (\bar{y}_p^d - \bar{y}^d)^2$$

$$MS_{Product*Subject}^d = \frac{R}{(P - 1)(N - 1)} \sum_{\substack{p=1 \\ n=1}}^{P, N} (\bar{y}_{pn}^d - \bar{y}_n^d - \bar{y}_p^d + \bar{y}^d)^2$$

121 Where $\bar{y}^d = \frac{1}{P} \sum_{p=1}^P \bar{y}_p^d$ and $\bar{y}_n^d = \frac{1}{N} \sum_{p=1}^P \bar{y}_{pn}^d$

122 Thus, a large $F_{Product}$ implies that differences between products are larger than the
 123 disagreement between subjects, and therefore allows considering the product means for a
 124 given descriptor as significantly different.

125 Let suppose that the product means are equal (H0), the so-called null hypothesis of the
 126 ANOVA. Then, under the usual assumptions of ANOVA (independence of observations,
 127 normality and variance homogeneity of residuals), F is compared to a theoretical Fisher
 128 distribution. The related p-value is calculated and corresponds to the probability to wrongly
 129 reject H0. If $p < 0.05$ (usual limit), one may conclude on significant differences between the
 130 product means (at least two product means are different).

131 3.2. Multidimensional discrimination

132 Multidimensional discrimination measures the separation of products in the sensory space
 133 generated by the descriptors relatively to subject disagreement. Multivariate Analysis of
 134 Variance (MANOVA), as an extension of ANOVA, allows to assess significance of this
 135 multidimensional discrimination and to quantify it.

136 Covariance matrices are the multivariate generalization of variances (cf. Appendix 2). If B
 137 (resp. W) denotes the product (resp. the interaction product*subject) covariance matrix
 138 between the descriptors, the discrimination matrix $W^{-1}B$ corresponds to the univariate
 139 $F_{Product}$. Successive eigenvectors of $W^{-1}B$ give linear combinations of descriptors
 140 optimizing the product discrimination measured by the corresponding eigenvalues. Specific
 141 functions of these eigenvalues offer multivariate tests of the product discrimination in the
 142 sensory space generated by the full set of descriptors (Appendix 2).

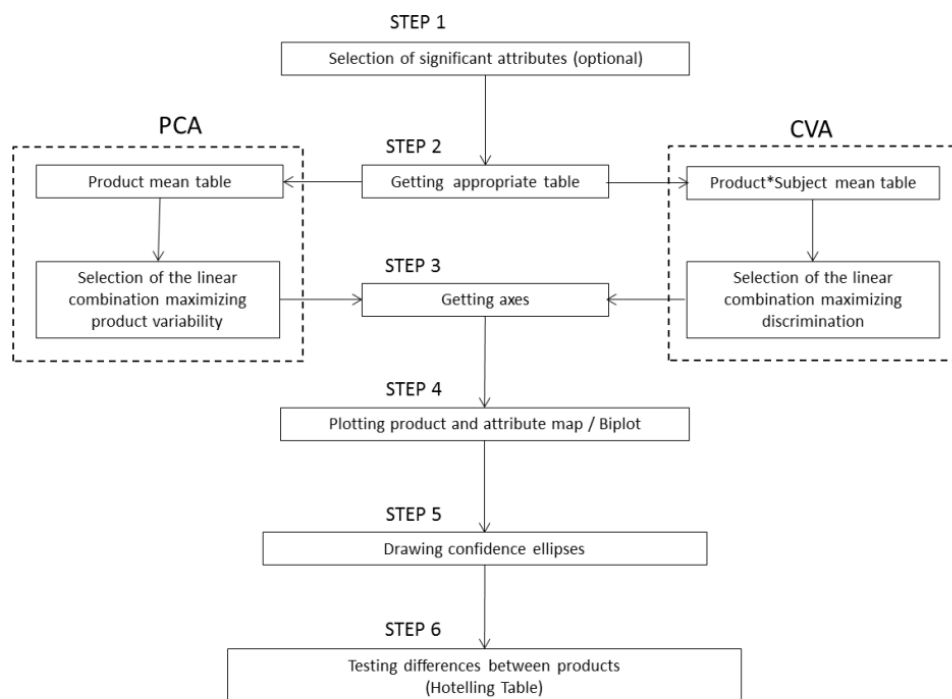
143 Note that the difference between multidimensional product variability and multidimensional
 144 product discrimination is the core of the difference between PCA and CVA. Indeed, while
 145 PCA extracts axes maximizing **variability between product means**, CVA axes maximize
 146 **product discrimination**.

147 **4. Obtaining a map**

148 **4.1. Overview of mapping methods**

149 PCA and CVA maps can be built following the steps presented in Figure 3. Each step will be
 150 developed and illustrated with the sample dataset.

151 FIG. 3. THE DIFFERENT STEPS TO PLOT A PCA OR CVA MAP



152 **Selection of significant descriptors**

154 In sensory profiling, several descriptors are scored, but, some of them can bring no
 155 differences among products and could consequently not be useful in multidimensional
 156 analysis. These descriptors are identified thanks to a two-way mixed model (cf. 3.1). The null
 157 hypothesis of the test of the ANOVA related to each descriptor is “All products have the same
 158 average”. When the p-value of this test is not significant ($p > 0.1$), the descriptor is not selected
 159 for the analysis.

160 This step is necessary before a correlation PCA, optional before a covariance PCA, and
 161 generally useless in CVA unless there are too much variables and/or colinearities. Note that in

162 correlation PCA, if one descriptor is not significant for the product effect, it would have the
 163 same importance than other descriptors in the map.

164 In the sample dataset, all descriptors had a significant product effect in the two-way model
 165 with random subject effect ($p < 0.1$), they were consequently included in the multidimensional
 166 analysis.

167 **4.2. Getting appropriate table**

168 PCA and CVA both include a preliminary step of centering of the descriptors. This step is
 169 necessary for the readability of the maps. Consequently, the average level of scores obtained
 170 by the descriptors is not a piece of information one can obtain from the PCA or CVA maps. If
 171 they are relevant, they must be identified with other methods (comparison of means, for
 172 example).

173 To sum up, plotting a map is based on:

- 174 - The centered product mean table for covariance PCA (presented in Table 1)
- 175 - The centered and standardized (divided by descriptor standard deviations) product
 176 mean table for correlation PCA. Standardization makes all descriptors having the same
 177 unit variance, and consequently the same importance in the analysis.
- 178 - The centered product*subject table for CVA

180 **TABLE 1: CENTERED PRODUCT MEAN TABLE FOR THE EXAMPLE DATASET**

	Butter	Crumbly	Dense	Elastic	Fat	Fruity	Hard	Intense	Milky	Salted	Smooth	Sticky
A	0.1587	1.1116	1.0259	1.0862	-0.6554	0.4186	1.3057	0.8616	0.1087	1.2381	-1.1248	-1.2345
B	-1.3483	2.1606	2.2967	1.8123	-1.4512	-1.2085	2.2557	-1.7998	-0.8694	-2.2036	-1.6342	-1.1491
C	0.8444	-1.0832	-1.0054	-0.9064	0.5915	0.6207	-1.1058	1.2606	0.5202	2.1464	0.7085	0.2498
D	-0.4462	0.5085	0.6707	0.5738	-0.2512	-0.4210	0.3880	-1.0946	-0.3955	-1.9245	-0.2592	-0.0887
E	0.8663	-1.5186	-1.5783	-1.8408	1.4207	0.9311	-1.7151	1.5335	0.4503	2.5536	1.3210	1.2081
F	0.2424	-1.1790	-1.4095	-0.7252	0.3457	-0.3408	-1.1286	-0.7613	0.1858	-1.8099	0.9887	1.01441

181 **4.3. Getting the axes**

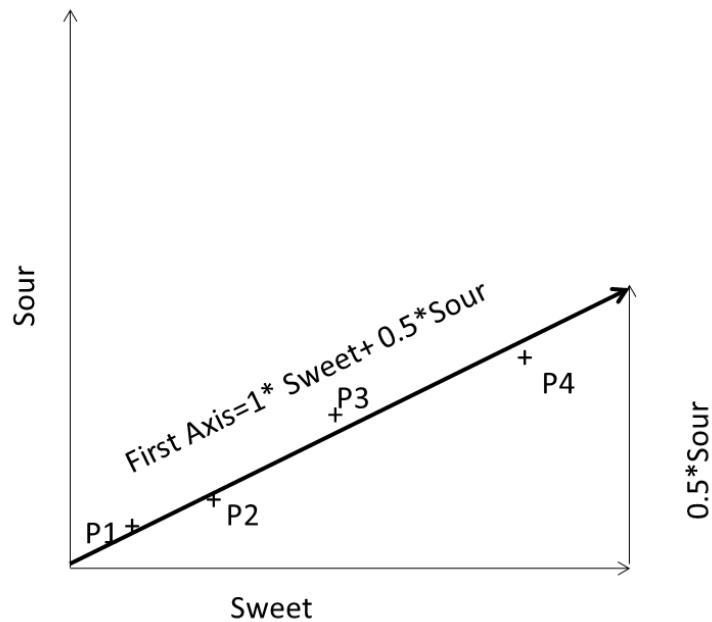
182 **4.3.1. What is an axis?**

183
 184 Suppose there are 2 descriptors only, then a map could naturally be drawn with first descriptor
 185 as X-axis and second descriptor as Y-axis and PCA or CVA map will just be a new choice of
 186 these axes but still staying in the same plan. Therefore, the map with two descriptors only
 187 would represent the totality of the data. However, in sensory analysis, more than 2 descriptors
 188 are generally used and the sensory space generated by them has D dimensions (with D greater
 189 than 2). Consequently, one map cannot represent the totality of the data anymore: the data has
 190 to be summarized. Therefore, axes representing most of the sensory differences among
 191 products have to be derived. They generate a subspace of lower dimensionality (usually 2
 192 dimensions). These “major” sensory dimensions are found thanks to relevant linear
 193 combinations of the descriptors. For example, in Figure 4, the axis represents the linear
 194 combination ($1 \cdot \text{Sweet} + 0.5 \cdot \text{Sour}$) or in vector form: $\begin{pmatrix} \text{Sweet} \\ \text{Sour} \end{pmatrix} \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$

195

196

FIG. 4. ILLUSTRATION OF A LINEAR COMBINATION IN A PLAN



197

198

199

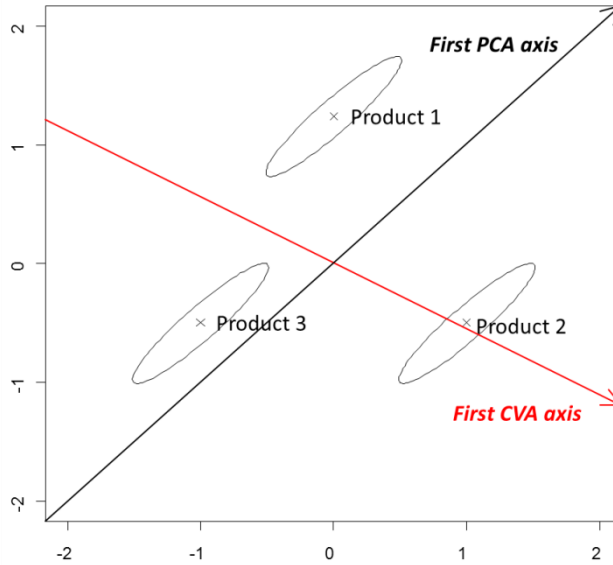
4.3.2. PCA and CVA axes characteristics

200 The PCA linear combinations maximize product **variability**, whereas the CVA ones
 201 maximize the product **discrimination**. Practically, matrix calculations based on the
 202 appropriate tables (on B for PCA, and on $W^{-1}B$ for CVA, cf. Section Multidimensional
 203 discrimination) are required to find these axes (cf. Appendix 2). Such axes are called
 204 Canonical Variates for CVA, and Principal Factors (or loadings) for PCA.

205 Figure 5 illustrates the differences between PCA and CVA first axis in a fictive example with
 206 2 descriptors and 3 products. Ellipses around product mean represent the distribution of the
 207 subject scores around this product. PCA finds the axis where product mean variability is
 208 maximal whereas CVA finds axes with a compromise between high product mean variability
 209 and small subject variability.

210

211 FIG. 5. DIFFERENCES BETWEEN THE FIRST AXIS OF PCA AND CVA



212 In the example dataset, the first two canonical variates are as following in Table 2. These
 213 canonical variates are “W-orthonormalized”. In other words, if v is a Canonical Variate,
 214 $v'Wv = 1$
 215

216 TABLE 2: CANONICAL VARIATES FOR THE EXAMPLE DATASET

	Can 1	Can 2
Butter	-0.0070	0.0038
Crumbly	0.0194	0.0217
Dense	-0.0017	-0.0058
Elastic	-0.0012	-0.0061
Fat	-0.0124	-0.0276
Fruity	0.0219	0.0255
Hard	0.0794	0.0586
Intense	-0.0379	0.0021
Milky	0.0154	-0.0312
Salted	-0.0524	0.0699
Smooth	0.0343	-0.0070
Sticky	-0.0180	-0.0292

217
 218 **4.3.3. Axes contribution**

219 As $W^{-1}B$ is the discrimination matrix of CVA, the contribution of an axis reflects the
 220 percentage of discrimination explained by this axis, denoting its relative importance for
 221 further interpretation

222 In PCA, the contribution of an axis represents the percentage of product variability explained
 223 by this axis.

224 **4.3.4. Number of axes to be analyzed**

225 A dimensionality test (Mardia, Kent and Bibbly 1989) allows finding how many axes are
 226 useful to interpret the data. The idea of the test is to keep only the dimensions of product
 227 space with information (and remove dimensions representing only noise). This test is based
 228 on the likelihood ratio criterion. Such a test does not exist in PCA. The number of axes is
 229 frequently selected by scree criterion and Kaiser criterion but does not result in statistical test.
 230 Consequently, the dimensionality test represents a real added value in favor of CVA.

231 However, we found over extensive use of this test that it tends to overestimate the number of
 232 axes that can further be sensorially interpreted. Thus, we see the information given by this test
 233 rather as an estimate of the maximum dimensionality of the discrimination space. Besides,
 234 this test can also be used as a tool for comparing dimensionality of different spaces, for
 235 instance several sensory panels having evaluated the same set of products with or without
 236 different sensory techniques.

237 **4.4. Representation of products and descriptors**

238 **4.4.1. Classical method**

239 Once axes are chosen, product and descriptor coordinates on these axes can be calculated.
 240 They are usually plotted on two different graphs: the product graph and the descriptor graph.
 241 In product graph, product means are projected on the chosen axes. In the example, the first
 242 coordinate of product A is obtained by calculating the scalar product of the centered mean
 243 scores of A and the first canonical axis:

$$(-0.1587) * (-0.0070) + \dots + (-0.0180) * -1.2345 = -0.0284$$

244 Thus, the first term of the sum corresponds to the product mean obtained for Butter times the
 245 coordinates of Butter for Can 1, and so on for each attribute.

246 The second coordinate (0.2450) is obtained by the same method, by projecting A scores on
 247 Can2 instead of Can1. Then, the product A will be represented by the point (-0.0284, 0.2450).

248 Regarding descriptor graph, the coordinates of each descriptor are calculated as following:
 249 The vector *Salted* containing all product scores for Salted is calculated. Then, product scores
 250 are projected on the first axis, giving the vector V_1 with P dimensions (its first coordinate here
 251 is -0.0284). The first coordinate of Salted arrow corresponds to the Pearson correlations
 252 between *Salted* and V_1 . The same type of calculation with the second axis gives the second
 253 coordinates of the extremity of the Salted arrow. Finally the salted coordinates are (-0.88,
 254 0.47). The reader can check these results on Figure 7.

255 **4.4.2. Biplot**

256 Plotting products and descriptors on the same graph is possible with a biplot. Biplot is a
 257 convenient way to interpret both product and descriptor relative locations simultaneously by

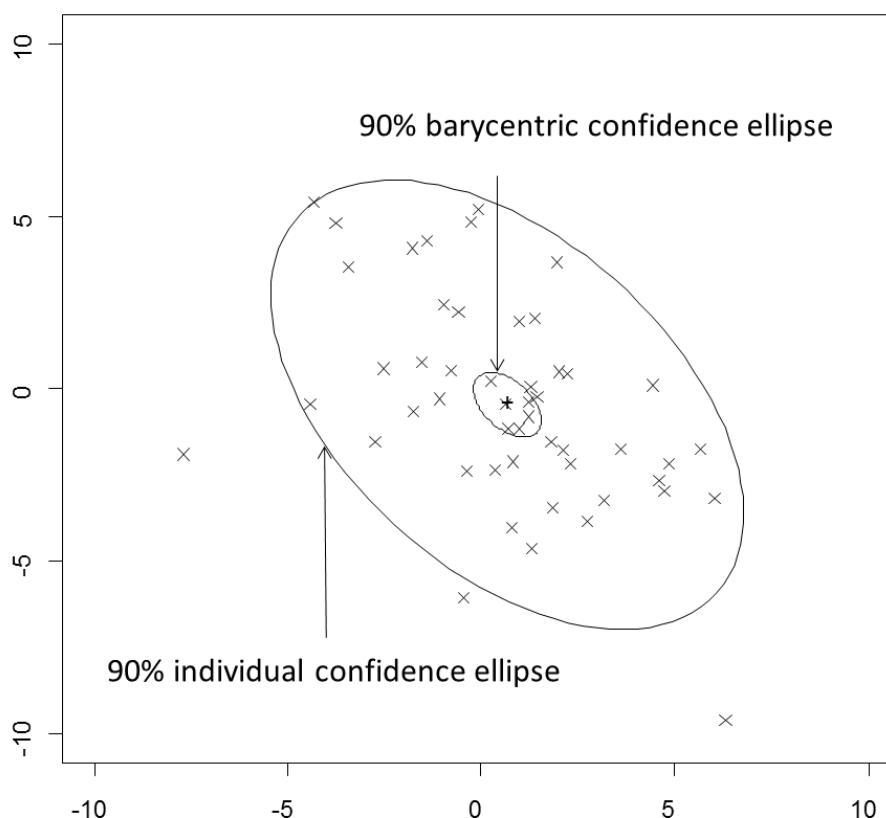
258 projecting products points on descriptors arrows. Therefore, such biplot is possible in PCA
 259 and CVA (Gower 2011) and is developed in Appendix 3. An example of biplot is presented in
 260 Figure 8.

261 **4.5. Subject representation: confidence ellipses of product means**

262 The distribution of subject scores around the product means can be visualized by confidence
 263 ellipses. In this purpose, for each product, subject scores are projected on the map (with the
 264 same method as the projection of products described in part 4.3.). Confidence ellipses are
 265 built with the matrix of covariance from each product scatterplot, supposing bi-normality of
 266 its distribution (Saporta 1986).

267 Ellipses can either represent a zone where the mean point has a probability of 90% to be
 268 (barycentric ellipses) or a zone where subject scores projections have a probability of 90% to
 269 be (subject ellipses). Barycentric ellipses can be interpreted as multidimensional confidence
 270 intervals of product means (Figure 6). The limit of 90% is chosen because of the small
 271 number of subjects in sensory profiling, but could also be 95 or 99%.

272 **FIG. 6. INDIVIDUAL AND BARYCENTRIC CONFIDENCE ELLIPSES WITH 50**
 273 **SUBJECTS**



274
 275 Ellipse intersections give also indications on differences between products. Intersected
 276 ellipses indicate that there is no difference between the related products, while separated
 277 ellipses indicate that related products are discriminated. Such ellipses can also be drawn on a
 278 PCA map with the same method (projecting subject scores as supplementary individuals).

279 Other methods to build confidence ellipses based on resampling exist but are not described in
 280 this paper.

281 **4.6. Representation of product pair differences: the Hotelling test**

282 As a reminder, a large F Product in MANOVA implies that at least 2 product are
 283 discriminated in the multidimensional product space. While confidence ellipses give an
 284 indication on product differences, the T^2 -Hotelling test is the appropriate statistical test to
 285 conclude on pairwise product differences. This test is the equivalent of the mono-dimensional
 286 t-test. Let k being the number of dimensions to be looked at, the T^2 -Hotelling test checks H_0
 287 “the 2 product mean vectors have the same location in the space generated by the first k axes”
 288 vs. H_1 “the 2 product mean vectors are not located at the same place in the space generated by
 289 the first k axes”. Then, a significant p-value indicates that the mean vectors are statistically
 290 different.

291 The number of tests can be high (about the squared number of products), consequently the
 292 limit for the p-value can be adjusted by a Bonferonni-type correction or any other procedure
 293 dealing with multiplicity.

294 If this test is based on the first two axes (k=2), results are most often similar to the indications
 295 given by ellipses intersections. In order to detect every multidimensional difference, one may
 296 advice to always run the Hotelling T^2 for k = total number of dimensions. However, the more
 297 dimensions, the more instable can be the test. If Hotelling T^2 test is done on all axes, too
 298 much importance might be given to the noise in the data: within variability will increase and
 299 so products could wrongly said to be equal. Consequently, the better solution to have an
 300 overview of the real multidimensional differences between products is to do the Hotelling test
 301 over only significant axes determined by the test given in section 4.4.4.

302 In the example, all p-values are lower than 0.05 except for the test between C and E (p=0.194)
 303 implying that C and E are multidimensionally equivalent. This table (Table 3) can be used to
 304 link similar products by segments on CVA map.

305 **TABLE 3: HOTELLING TABLE FOR THE EXAMPLE DATASET**

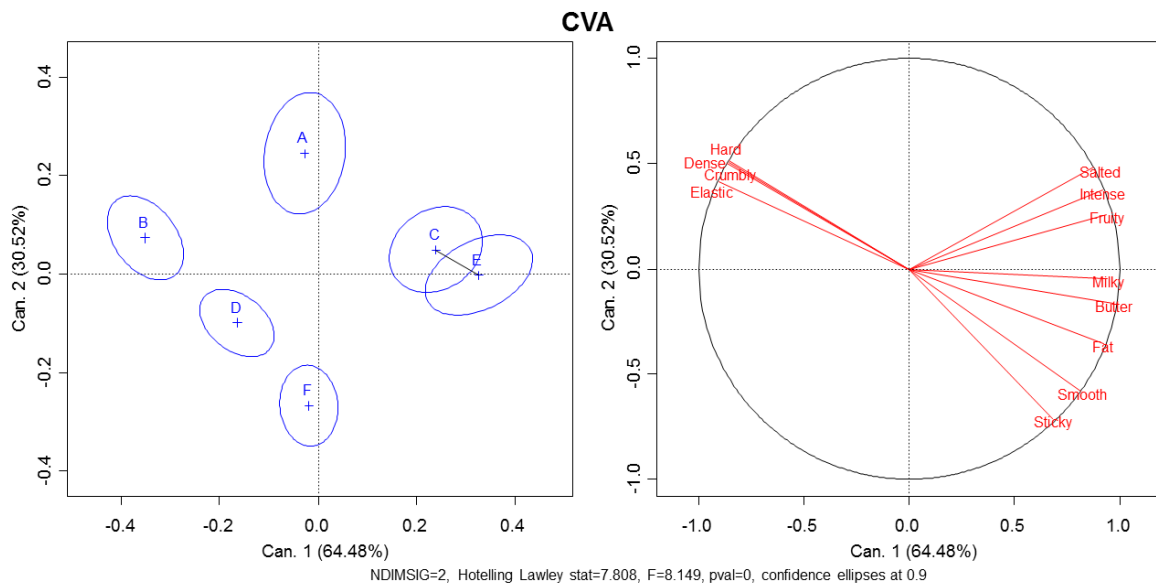
	A	B	C	D	E	F
A	1	0.00	0.00	0.00	0.00	0.00
B	0.00	1	0.00	0.00	0.00	0.00
C	0.00	0.00	1	0.00	0.19	0.00
D	0.00	0.00	0.00	1	0.00	0.00
E	0.00	0.00	0.19	0.00	1	0.00
F	0.00	0.00	0.00	0.00	0.00	1

306
 307 **4.7. Interpretation of a CVA map**

308 Figure 7 illustrates the CVA plot (without biplot) for the example dataset.

309

FIG. 7. CVA MAP OF THE EXAMPLE DATASET



310

311 The first plan accounts for 94.99% of the product discrimination (64.48% explained by the
 312 first axis). The footnote shows that products are significantly discriminated as exemplified by
 313 the Hotelling-Lawley statistic approximated by an F-statistic of 8.12 corresponding to a p-
 314 value close to 0. Furthermore, the product space can be resumed in a 2 dimension subspace
 315 (NDIM=2). Confidence interval limit for ellipses and Hotelling test is set to 0.9 (90%).

316 The Hotelling test discriminates all pairs of products but C and E which are connected with a
 317 segment on the map in order to show their similarity. C and E are more Salty/Intense/Fruity
 318 and less Hard/Dense/Crumbly/ Elastic than the other products.

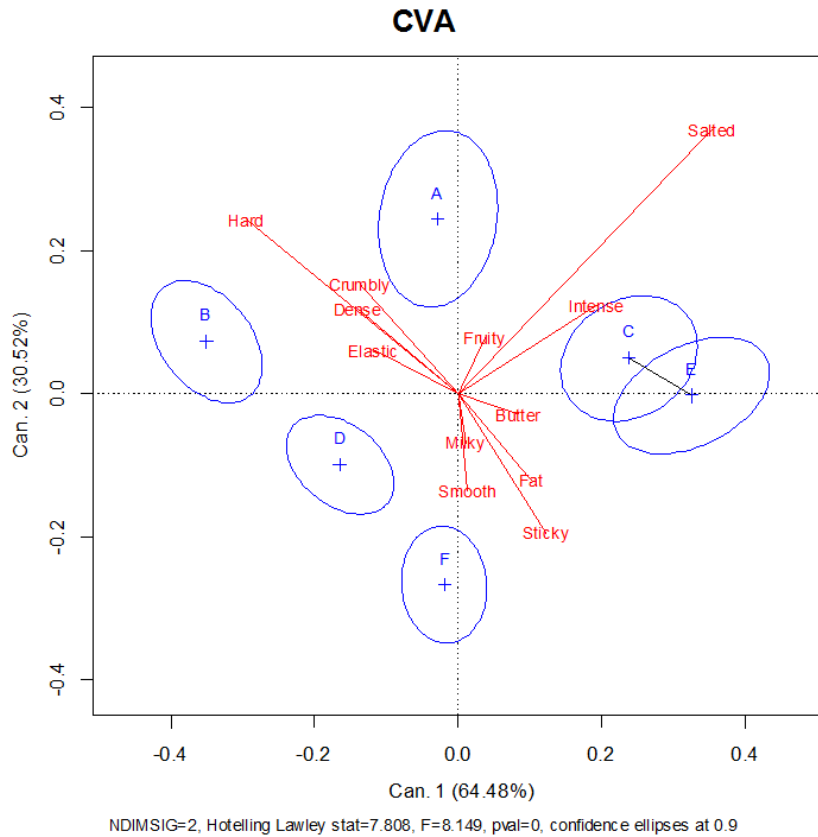
319 Biplot (Figure 8) shows the same product locations, but different descriptor ones.

320

321

322

FIG 8. CVA BIPLLOT OF THE EXAMPLE DATASET



323

324 The size of a descriptor arrow is proportional to the quality of the descriptor representation on
 325 the 1-2 plot. For example, Hard and Salted seem to be well represented on this plot.

326 Consequently, they have an important role in the multidimensional discrimination of the
 327 products. On the contrary, Butter, Fruity and Milky are not well represented because they are
 328 less powerful in discriminating the products. These results are confirmed by the
 329 unidimensional tests presented on table 4.

330

TABLE 4: PRODUCT DISCRIMINATION FOR THE EXAMPLE DATASET

	F	Pvalue
Butter	11.63	<0.001
Crumbly	25.31	<0.001
Dense	27.61	<0.001
Elastic	16.45	<0.001
Fat	13.95	<0.001
Fruity	9.76	<0.001
Hard	46.97	<0.001
Intense	27.37	<0.001
Milky	7.14	<0.001
Salted	53.56	<0.001
Smooth	18.32	<0.001
Sticky	13.17	<0.001

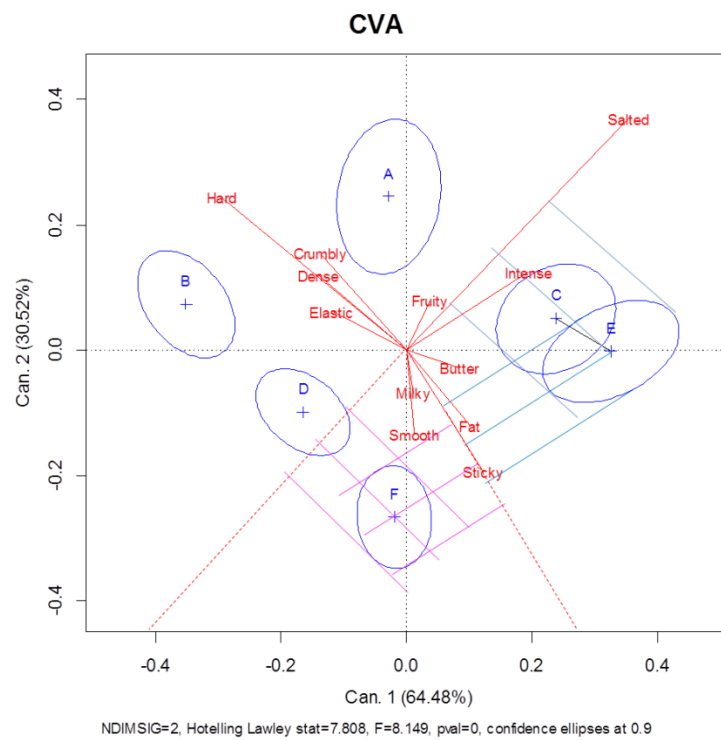
331

332 Furthermore, the orthogonal projection of a product on a descriptor arrow added to the grand
 333 mean of this descriptor approximates as much as possible the real mean score of this product.

334 The projection of ellipses on descriptors can also give indications about their differences
 335 (Figure 9). For example, E and F are different on Salty ($p=0.00$ by regular t-test), but are not
 336 different on Sticky ($p=0.72$ by t-test).

337 However this indication is not always exact: for Fat, the p-value of the t-test between E and F
 338 is 0.03 while the ellipse projections do intersect. Consequently, the map is not sufficient to
 339 conclude to differences between products, statistical tests should be done to validate the visual
 340 results.

341 FIG. 9. PROJECTION OF PRODUCT ELLIPSES ON BILOT ARROWS



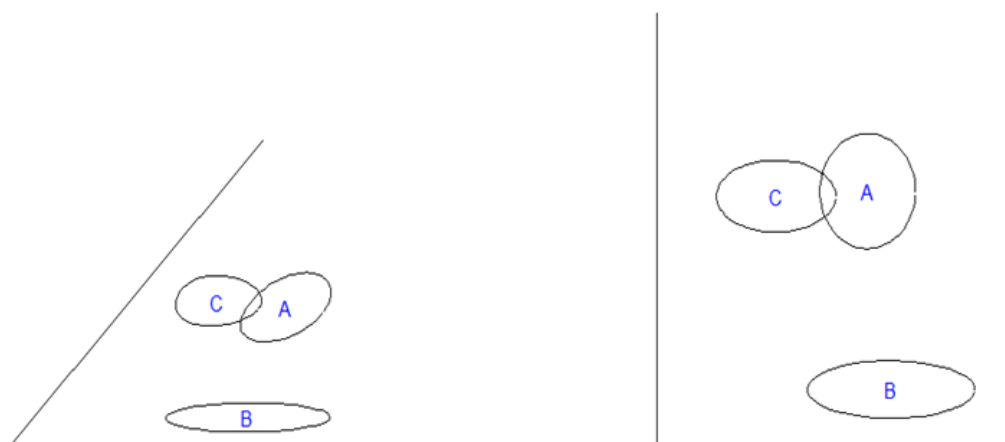
342

343 **5. Discussion**

344 **5.1. Non-orthogonality of CVA axes**

345 While PCA axes are orthogonal (not correlated), it is not the case of CVA axes which are just
 346 W-orthonormal. That means that axes are not systematically right-angled. However, plotting
 347 them orthogonally is commonly used in Discriminant Analysis since correcting the angle does
 348 not change any aspect of interpretation of the map, but makes the plot less readable (Figure
 349 10).

350 FIG. 10. ILLUSTRATION OF THE FACT THAT PLOTTING CANONICAL
 351 VARIATES ORTHOGONAL DOES NOT CHANGE DIAGNOSIS



352
 353 **5.2. Consequences of matrix inversion**

354 The major problem in CVA is that it uses a matrix inversion (cf. Appendix 2). Indeed, when
 355 there are strong colinearities between the vectors of W , inversion can be unstable (the
 356 determinant of the matrix W is close to 0). The condition index (CI) measures the presence (or
 357 absence) of colinearities issues (Callaghan 2008).

- 358 - Weak conditioning issue ($CI < 10$)
- 359 - Moderate to strong conditioning issue ($10 < CI < 30$)
- 360 - Severe conditioning issue ($CI > 30$)

361 Some bad conditioning can easily happen in case of large dimension of W (too many
 362 descriptors). The diagnosis is produced in the proposed CVA-package: for the example
 363 dataset CI equals 5.65.

364 Hotelling test also uses a matrix inversion and thus may sometimes be not computable.

365 **5.3. CVA: a map based on a statistical model**

366 CVA presents the advantage to rely on a statistical model. Our work has been done on the
 367 classical two-way (product and subject) model with subject considered as a random effect, but
 368 it can be extended to other models, including other effects. Furthermore, Fisher Discriminant
 369 Analysis is nothing else but a CVA with a one-way model. Consequently, performing a CVA
 370 with a program for Discriminant Analysis is possible with some data transformations: scores
 371 have to be averaged by replicate and centered by subject (Appendix 4).

372 **6. Conclusion**

373 Whereas product map in sensory studies are usually produced by a PCA of the product mean
 374 scores, it is CVA which allows plotting a map in accordance to the classical ANOVA model
 375 taking subject variability into account. This paper has highlighted the relevance of this
 376 method for the analysis of descriptive sensory data, explained its principle and offered an R-
 377 package to use it (freely available on www.timesens.com). In addition to this method,

378 multidimensional tests of differences between products are provided. Interpreting
 379 multidimensional differences with these tests is advised instead of interpreting intersection
 380 between confidence ellipses. Thus, errors due to wrong visual conclusions could be avoided.

381 **References:**

- 382 CALLAGHAN, K., CHEN, J., 2008. Revisiting the collinear data problem: an assessment of estimator
 383 'ill-conditioning' in linear regression. *Practical Assessment, Research And Evaluation* 13.
 384 GOWER, J.L., S.; LE ROUX, N., 2011. Understanding biplot. (JOHN WILEY AND SONS, L., ed.,
 385 eds.).
 386 GRUNG, B. and MANNE, R. 1998. Missing values in principal component analysis. *Chemometrics*
 387 *and Intelligent Laboratory Systems* 42, 125-139.
 388 HUSSON, F. and JOSSE, J. 2013. Handling missing values in multiple factor analysis. *Food Quality*
 389 *and Preference* 30, 77-85.
 390 MARDIA, K.V., KENT, J.T. and BIBBLY, J.M. 1989. *Multivariate Analysis*.
 391 NÆS, T. and LANGSRUD, Ø. 1998. Fixed or random assessors in sensory profiling? *Food Quality*
 392 *and Preference* 9, 145-152.
 393 NOBLE, A.C.W., A. A.; LANGRON, S.P. 1984. Descriptive analysis and quality ratings of 1976
 394 wines from four Bordeaux communes. *Journal of the Science of Food and Agriculture* 35,
 395 88-98.
 396 PELTIER, C., BROCKHOFF, P.B., VISALLI, M. and SCHLICH, P. 2014. The MAM-CAP table: A
 397 new tool for monitoring panel performances. *Food Quality and Preference* 32, Part A, 24-27.
 398 PORCHEROT, C. and SCHLICH, P. 2000. FLASH table and canonical mapping of potato varieties.
 399 *Food Quality and Preference* 11, 163-165.
 400 SAPORTA, G., HATABIAN, G., ed. 1986. *Régions de confiance en analyse factorielle*, Elsevier
 401 Science Publisher B.V.

402

403

404

405 **Appendix 1: Using CVA-package**

406 The *CVApack* R-package computes a CVA and returns most of results and figures presented
 407 in this paper. It is available at www.timesens.com as a zip file. This zip file must be recorded
 408 on the computer and then installed in R from this zip file (Packages/Install from zip files).

409 Then, the package is loaded by the command

410 `library(CVApack)`

411 and information about the package can be obtained by typing `?CVApack`. The example dataset
 412 is included in the package and can be loaded with:

413 `data(cheeses)`

414 To obtain CVA results with two graphs, enter the commands:

415 `resCVA=CVA(extendedData=cheeses,representation="twoMaps")`

416 where `extendedData` corresponds to a table containing the data whose colnames are
 417 “ProductCode”, “SubjectCode”, “Replicate”, “Session”, first attribute, ... , last attribute.

418 Then, to plot the map, enter

419 `plotCVA(resCVA)`

420 To plot the CVA with the biplot, the following commands should be entered:

421 `resCVA=CVA(extendedData=data, representation="biplot")`

422 `plotCVA(resCVA)`

423 Other options are available to build individual or barycentric ellipses, to choose the limit for
 424 the confidence ellipses. Help is available by typing `help("CVA")` or `?CVA`. As outputs of
 425 `resCVA`, `B`, `W`, the eigen-vectors and values of $W^{-1}B$, the coordinates of products and
 426 variables, the conditioning index, the number of dimensions and the Hotelling Table are
 427 presented in a list. The TimeSens© software features a friendly interface between the R
 428 library and the user.

429 **Appendix 2: Theoretical background for CVA**

430 **1- Writing the covariance matrices of product and interaction product*subject**

431 Let us denote PROD (resp. INTER) as the P*D (resp. P.N*D) matrix of product (resp.
 432 interaction) effects:

$$PROD = \begin{bmatrix} (\bar{y}_{1..}^1 - \bar{y}_{...}^1) & \cdots & (\bar{y}_{1..}^D - \bar{y}_{...}^D) \\ \vdots & \ddots & \vdots \\ (\bar{y}_{P..}^1 - \bar{y}_{...}^1) & \cdots & (\bar{y}_{P..}^D - \bar{y}_{...}^D) \end{bmatrix} = (\bar{y}_{p..}^d - \bar{y}_{...}^d)_{p=1..P, d=1..D}$$

$$INTER = (\bar{y}_{pn.}^d - \bar{y}_{p..}^d - \bar{y}_{.n.}^d + \bar{y}_{...}^d)_{p=1..P, n=1..N, d=1..D}$$

433 Where $\bar{y}_{...}^d$ is the average of the scores from descriptor d over all products; $\bar{y}_{p..}^d$ the score
 434 average from descriptor d and product p; $\bar{y}_{.n.}^d$ the score average from descriptor d and subject
 435 n; and $\bar{y}_{pn.}^d$ the score average from descriptor d, subject n and product p.

436 Then, let be $B = \frac{1}{(P-1)} (PROD)'(PROD)$ (resp. $W = \frac{1}{(P-1)(N-1)} (INTER)'(INTER)$) the
 437 d*d matrices of covariance of the product (resp. interaction) effect. Note that B (resp. W)
 438 diagonal coefficients are the sum of squares of the product (resp. product*subject interaction)
 439 effect for the corresponding descriptor.

$$B[d, d] = SS_{product}^d = R \sum_{p=1}^P (\bar{y}_{p..}^d - \bar{y}_{...}^d)^2$$

$$W[d, d] = SS_{product*subject}^d = R \sum_{p=1, n=1}^{P, N} (\bar{y}_{pn.}^d - \bar{y}_{p..}^d - \bar{y}_{.n.}^d + \bar{y}_{...}^d)^2$$

440 2- Diagonalization of the discrimination matrix

441 To find the linear combination which maximizes the product effect over the interaction effect
 442 leads to find x maximizing the fraction:

$$\frac{\|PROD * x\|^2}{\|INTER * x\|^2} = \frac{x^T B x}{x^T W x} \quad (1)$$

443 It means that we can search $y = W^{1/2} x$ such as $\frac{y^T W^{-1/2} B W^{-1/2} y}{y^T y}$ be maximal. $W^{-1/2} B W^{-1/2}$
 444 being a symmetric matrix, this optimization problem has well-known solutions: the
 445 eigenvectors of $W^{-1/2} B W^{-1/2}$. Consequently, the solutions y fills the following conditions,
 446 where λ is an eigenvalue of $W^{-1/2} B W^{-1/2}$.

$$447 \quad W^{-1/2} B W^{-1/2} y = \lambda y \text{ and } \|y\|^2 = 1$$

$$448 \quad \Leftrightarrow W^{-1/2} B x = \lambda W^{1/2} x \text{ and } \|W^{1/2} x\|^2 = 1$$

$$449 \quad \Leftrightarrow W^{-1} B x = \lambda x \text{ and } x^T W x = 1$$

450 Consequently, the vectors maximizing (1) are the eigenvectors of $W^{-1} B$ so-called
 451 **discrimination matrix**. They are obtained by the **diagonalization** of this matrix which allows
 452 to find eigenvalues $(\lambda_1, \dots, \lambda_D)$ and eigenvectors $v = (v_1, \dots, v_D)$ such as $W^{-1} B v_i = \lambda_i v_i$ for
 453 $i= 1 \dots \min(D, P-1)$.

454 Thus, $W^{-1} B$ can be written $(v_1; \dots v_D)^T \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_D \end{pmatrix} (v_1; \dots v_D)$

455 $v_1; \dots v_D$ are chosen as axes of the CVA map, are W -orthonormalized ($x^T W x = 1$) and
 456 maximize the discrimination of the products.

457 It should be noticed that if descriptors were uncorrelated, B and W would have been diagonal,
 458 as the $W^{-1}B$ matrix, and therefore the $W^{-1}B$ diagonal coefficients would have been the F
 459 values of the univariate ANOVAs by descriptor.

460 Note that in usual PCA of sensory data, it is the B matrix which is diagonalized.

461 3- Calculating the discrimination p value

462 The ratio of one eigenvalue with the sum of eigenvalues corresponds to the discriminative
 463 power of the corresponding axis. In MANOVA, the eigenvalues of $W^{-1}B$ are used to
 464 calculate a multivariate F which allows to test for multidimensional product discrimination by
 465 several methods: Roy, Hotelling-Lawley, Wilk's lambda, Pillai, whose statistics are noted
 466 Λ_{Wilks} , $Tr_{Hotelling}$, R , Tr_{Pillai} . Hotelling-Lawley trace is our default statistic.

$$R = \max_{d=1..D} \lambda_d$$

$$Tr_{Hotelling} = \sum_{d=1}^D \lambda_d$$

$$\Lambda_{Wilks} = \prod_{d=1}^D 1/(1 + \lambda_d)$$

$$Tr_{Pillai} = \sum_{d=1}^D \lambda_d/(1 + \lambda_d)$$

467 Then, a function of each of these statistics can be approximated, possibly asymptotically, by a
 468 Fisher distribution and consequently used to calculate a p-value allowing assessing if at least
 469 two products are significantly located in different areas of the sensory space.

470 Appendix 3: Biplot in PCA and CVA

471 The principle of a biplot is to find a map where product projections and descriptor arrows can
 472 directly be compared. Therefore, the orthogonal projection of one product onto a descriptor
 473 arrow should approximate the average (up to a multiplicative constant) obtained by this
 474 product for this descriptor. A Singular Value Decomposition (SVD) is used for finding
 475 product and descriptor coordinates.

476 In PCA, let X be the column-centered product mean matrix of size $P \times D$. Then, the SVD
 477 decomposition of X is:

$$478 \quad X = U \Sigma V^T$$

479 Where U is an orthonormal $P \times P$ matrix ($U^T U = Id$), V is an orthonormal $D \times D$ matrix and Σ a
 480 $P \times D$ matrix whose diagonal coefficients are positive or null and all others are null.

481 Then, the so-called “distance biplot” consists in choosing $U\Sigma$ to represent products and V to
 482 represent descriptors. As $B = X^T X = V\Sigma U^T U\Sigma V^T = V\Sigma^T \Sigma V^T$, this decomposition allows
 483 finding again the diagonalization of B with V as eigenvectors. Note that product coordinates
 484 equal XV in PCA, and according to the SVD decomposition of X , it is also equal to $U\Sigma$.
 485 Consequently, product coordinates are the same in distance biplot and classical PCA.

486 Regarding the descriptors, the norm of an i^{th} descriptor in 1-2 map is $\sum_{j=1}^2 v_{ij}^2$ by noting
 487 v_{ij} the i^{th} component of the j^{th} eigenvector v_j .

488 Moreover, as v is an eigenvector, its norm (over all dimensions) equals 1.

$$\|v_j\|^2 = \sum_{i=1}^D v_{ij}^2 = 1$$

489 Consequently, $\sum_{i=1, j=1}^{D, 2} v_{ij}^2 = 2$ and the sum of the norms of all descriptors in 1-2 map
 490 should equal 2. The size of the represented arrow reflects the relative importance of the
 491 descriptors in this map: it gives indications about the product **variability** explained by this
 492 descriptor.

493 Regarding the biplot in CVA, let L be a $D \times D$ matrix such as $LL^T = W^{-1}$ with L a symmetric
 494 matrix and consider XL as the matrix of the centered scores weighted by the disagreement
 495 within subjects. The SVD decomposition of this matrix gives U and V such as $XL = U\Sigma V^T$.
 496 Product coordinates equal $U\Sigma$ whereas descriptors coordinates equal V .
 497 As in PCA, the size of arrows indicates the product **discrimination** explained by the
 498 descriptor. For this reason, this biplot is not exactly the same biplot than the biplot presented
 499 in Gower (2011).

500 **Appendix 4: From CVA to LDA**

501 With Appendix 2 notations, the classical two-way model implies a decomposition of scores
 502 as:

$$y_{pnk}^d = \bar{y}_{...}^d + (\bar{y}_{p..}^d - \bar{y}_{...}^d) + (\bar{y}_{.n.}^d - \bar{y}_{...}^d) + (\bar{y}_{pn.}^d - \bar{y}_{p..}^d - \bar{y}_{.n.}^d + \bar{y}_{...}^d) + (y_{pnk}^d - \bar{y}_{pn.}^d)$$

503 Let z_{pn}^d be the score related to y_{pnk}^d such as:

$$504 \quad z_{pn}^d = y_{pnk}^d - (\bar{y}_{.n.}^d - \bar{y}_{...}^d) - (y_{pnk}^d - \bar{y}_{pn.}^d)$$

505 Then, z_{pn}^d correspond to the score averaged over replicates and centered by subjects and is
 506 decomposed as following in the one-way product model.

$$507 \quad z_{pn}^d = \bar{z}_{...}^d + (\bar{z}_{p..}^d - \bar{z}_{...}^d) + (\bar{z}_{pn.}^d - \bar{z}_{p..}^d)$$

508 Given that

$$(\bar{z}_{p..}^d - \bar{z}_{...}^d) = y_{p..}^d - y_{...}^d$$

509 and

$$z_{pn}^d - \bar{z}_{p..}^d = y_{pn.}^d - (\bar{y}_{.n.}^d - \bar{y}_{...}^d) - y_{p..}^d = \bar{y}_{pn.}^d - \bar{y}_{p..}^d - \bar{y}_{.n.}^d + \bar{y}_{...}^d$$

510 The product term in the one way model with z_{pn}^d correspond to the product term in the
 511 classical model, and the error term in the one way model corresponds to the interaction term
 512 for the two-way model with y_{pnk}^d . Consequently, Linear Discriminant Analysis (LDA) on the
 513 data averaged over replicates and centered by subject corresponds to the CVA map. However,
 514 the results of the MANOVA test in LDA will be different and wrong, due to a loss of degrees
 515 of freedom brought by averaging and centering. Consequently, software calculating LDA can
 516 be used to obtain CVA maps, but does not give the appropriate MANOVA statistics.

517

1 **Multiplicative decomposition of the scaling effect in the Mixed Assessor Model into a**
2 **descriptor-specific and an overall coefficients**

3 C. Peltier^{1,2,3,*}, M. Visalli^{1,2,3}, P. Schlich^{1,2,3}

4 ¹ CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France.

5 ² INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France.

6 ³ Université de Bourgogne, UMR Centre des Sciences du Goût et de l'Alimentation, F-21000
7 Dijon, France.

8
9 * caroline.peltier@dijon.inra.fr

10 **Abstract**

11 In 1994, Brockhoff and Skovgaard presented the so-called Assessor Model, including a
12 “scaling coefficient” expressing the predisposition of a panelist to spread more or less his
13 scores than the panel on a given sensory descriptor. This paper discusses potential scaling
14 causes, and then proposes a decomposition of the scaling coefficient into two components:
15 (i) an overall scaling coefficient, independent of the descriptors, expressing a psychological
16 trend of the panelist towards the scoring task in general; (ii) a corrected scaling coefficient
17 for each descriptor, expressing specific sensitivity of the panelist to the descriptor.

18 Applied to 187 sensory datasets, this decomposition showed that panelists over sixty years
19 old and smokers tend to spread less their scores than other panelists regardless of the
20 sensory modalities; further, the corrected scaling coefficients were more heterogeneous
21 among panelists for aroma and flavor compared to texture and appearance sensory
22 modalities.

23 **Keywords:** scaling; Mixed Assessor Model; sensometrics; SensoBase; meta-analysis

24 **Highlights:**

- 25 • Scaling effect can be decomposed into overall and corrected scaling
- 26 • Overall scaling can represent a psychological scaling effect
- 27 • Corrected scaling can represent a physiological scaling effect
- 28 • Over 60 years old panelists and smokers spread less their scores than other panelists
- 29 • Negativity issue of scaling disappears with overall scaling

30
31 1. Introduction:

32 Sensory panel data, where panelists score different products on a continuous scale, are
33 often blurred by different factors.

34 According to Meilgaard, Gail Vance Civille, and Carr (1999), factors influencing the scores can
35 be physiological (such as adaptation, enhancement, poor physical conditions...) or
36 psychological (such as halo effect, lack of motivation, timidity...). Therefore, two people's

37 verdicts can be caused either by a difference in their sensorial sensitivity or by a difference in
38 their mental treatment of the sensations.

39 Lim (2011) decomposes the sensory process in two major steps. The “physiological step”
40 consists of a sensory transduction of the stimulus into an internal representation including
41 sensation quality, intensity and hedonic values. This step can be influenced by the sensory
42 perceptual context. The “psychological step” includes evaluative and decision making
43 processes. This step can be modified by the response context, including the instruction given
44 to the panelists.

45 Physiological individual characteristics can affect the scores. Thus, smoking could affect taste
46 (Krut, 1961) and odor (Ahlström, Berglund, Berglund, Engen, & Lindvall, 1987). Aging could
47 decrease the sensitivity for sweet and taste (Booth, Kohrs, & Kamath, 1982) or the ability to
48 recognize all four tastes (Booth et al., 1982). Another example illustrating the physiological
49 differences is the individual sensitivity to bitterness (Meilgaard et al., 1999).

50 As an example of differences in individual psychology, it is well known that some people
51 tend to use the extremes of any scale, thereby exerting more than their share of influence
52 over the panel’s results, while others tend to stick to the central part of the scale and to
53 minimize differences between samples. This effect is also described as “Error of Central
54 Tendency” (Herbert & L., 2004) or as a bias named “Timidity versus Capriciousness”
55 Meilgaard et al. (1999).

56 The scaling effect was introduced in 1994 (P. M. Brockhoff & Skovgaard, 1994) thanks to the
57 so-called Assessor Model, in order to correct the Error of Central Tendency. However, this
58 model did not allow testing product effect and was consequently not really used by sensory
59 scientists. P. B. Brockhoff, Schlich, and Skovgaard (2015) filled this gap with the Mixed
60 Assessor Model (MAM) which decomposes the classical disagreement term in the two-way
61 mixed model into a scaling term and a pure disagreement term. The scaling term is
62 expressed by a “scaling coefficient” measuring the predisposition of a panelist to spread his
63 scores more than the panel. The MAM was used to calculate and describe panel and panelist
64 performances in the MAM-CAP table (Peltier, Brockhoff, Visalli, & Schlich, 2014).

65 Until now, the MAM scaling coefficient was applied for a single descriptor at a time, whereas
66 sensory profiling is typically a multiple-descriptor task. Romano et al. (2015) proposes a
67 multivariate assessor model for the multivariate analysis of the scaling effect. This approach
68 based on PCA still considers one scaling coefficient by descriptor and subject. Our objective
69 is to have only one scaling coefficient by subject. Indeed, dealing with different descriptors
70 can give indications about possible causes of scaling. If the MAM scaling coefficients were
71 similar for all descriptors, it could testify a specific behavior due to a “psychological” effect
72 (self-confidence, shyness ...). On the contrary, different MAM scaling coefficients could
73 represent a physiological effect (hypo- or hyper- sensitivity) specific for each descriptor.

74 Consequently, this paper decomposes the classical scaling into a component specific to each
 75 descriptor (“corrected scaling”) and a component common to all descriptors (“overall
 76 scaling”). This decomposition was studied on 187 datasets. In this meta-analysis, the impact
 77 of panelist characteristics (age, gender, smoking status) on overall scaling was studied. The
 78 stability of the overall scaling was analyzed based on panelists being present in at least 4
 79 datasets. Finally, corrected scaling was investigated in order to find out whether the sensory
 80 modality impacts the heterogeneity of corrected scaling among panelists.

81 2. Material and Methods:

82 2.1. Decomposition of MAM scaling coefficient

83 2.1.1. Calculation of MAM scaling coefficient

84 Replicated quantitative descriptive sensory data, where J products are scored by I assessors
 85 in K replications on D descriptors can be analyzed with the MAM model presented in P. B.
 86 Brockhoff et al. (2015):

$$y_{ijk}^d = \mu^d + \alpha_i^d + v_j^d + \tilde{b}_i^d x_j^d + \delta_{ij}^d + \varepsilon_{ijk}^d$$

87 where, for descriptor $d = 1 \dots D$, μ^d is the general average, α_i^d the panelist effect, v_j^d the
 88 product effect, $j = 1 \dots J$, $x_j^d = y_{j.}^d - y_{...}^d$ the centered product averages inserted as a
 89 covariate, δ_{ij}^d the pure disagreement term and $\varepsilon_{ijk}^d, k = 1 \dots K$, the error term.

90 $\tilde{b}_i^d, i = 1 \dots I$, corresponds to $b_i^d - 1$ where b_i^d is the slope of the linear regression of the
 91 centered panelist scores $y_{ij.}^d - y_{i..}^d$ on the centered panel scores $y_{j.}^d - y_{...}^d$ (Figure 1).

92 b_i^d is called the **MAM scaling coefficient** in this paper. This coefficient allows detecting
 93 whether a panelist spreads more ($b_i^d > 1$) or less his scores ($0 < b_i^d < 1$) than the panel.

94 [Insert Figure 1]

95 Note that this coefficient can also be negative for three main reasons:

- 96 - The panelist really feels the descriptor inversely from the rest of the panel. In this
 97 case, the negativity of the scaling coefficient expresses a disagreement with the
 98 panel which should not be corrected.
- 99 - The products are not well discriminated. In this case, on a continuous scale, the order
 100 of the product means given by the panel is random and can consequently be
 101 opposite to the order of the product means given by the panelist.
- 102 - The panelist misunderstood the descriptor or the labels of the scale and scores at the
 103 opposite of what he feels: a correction should be done. It could happen during
 104 training sessions.

105 In a study based on 369 datasets, at least one negative b_i^d was observed for two thirds of
 106 the descriptors (P. B. Brockhoff et al., 2015). As a real explanation is difficult to determine, it

107 is recommended not to apply MAM to the data related to negativity issues (P. B. Brockhoff
108 et al., 2015).

109 2.1.2. Definition of “Overall scaling coefficient”

110 The computation of the MAM scaling coefficient presented in the previous section extended
111 to all descriptor scores simultaneously provides a unique scaling coefficient by panelist
112 named the “overall scaling coefficient” β_i . First, panelist and panel scores were centered by
113 descriptor. Then, β_i is obtained by calculating the slope of the regression of panelist scores
114 on panel scores all descriptors confounded as presented in Figure 2.

115 [Insert Figure 2]

116 When $\beta_i > 1$, the panelist tends to systematically spread his scores more than the panel.
117 When $0 < \beta_i < 1$, the panelist tends to systematically spread his scores less than the panel.
118 As all descriptors are involved, this coefficient could represent either a psychological effect
119 as presented in Meilgaard et al. (1999) or a hyper- (hypo-) sensitivity for most of the
120 descriptors in the dataset, or a combination of both.

121 2.1.3. Calculation of “Corrected scaling coefficient”

122 2.1.3.1. Interest of correcting the MAM scaling

123 Let’s assume that a given panelist has an overall scaling of 3. If he has a MAM scaling
124 coefficient of 1 for bitterness, then, the classical interpretation indicates neither
125 hyposensitivity nor hypersensitivity. However, as this panelist is used to spread his scores 3
126 times larger than the panel ($\beta_i = 3$), the fact that he uses the same scale than the panel
127 only for bitterness reflects hyposensitivity towards this attribute. The overall scaling
128 coefficient should consequently be taken into account in the detection of hypo- or hyper-
129 sensitivity.

130 2.1.3.2. Obtaining the “corrected scaling coefficient”

131 The “corrected scaling coefficient” $\gamma_{i,d}$ aims to detect hypo- or hyper- sensitivity freed from
132 a potential psychological scaling effect. It corresponds to the MAM scaling coefficient
133 divided by the overall scaling coefficient. A decomposition of MAM scaling is consequently
134 obtained as following:

$$b_{i,d} = \beta_i * \gamma_{i,d}$$

135 A hyper-sensitivity (or hypo- sensitivity) relative to the panel is detected when the corrected
136 scaling coefficient is higher (or lower) than 1. Therefore, the corrected scaling of bitterness
137 for the panelist of the previous example is $\gamma_{i,d} = 1/3$, denoting a possible hypo-sensitivity
138 relative to the panel for bitterness. Note that when a panelist has no overall scaling ($\beta_i =$
139 1), the corrected scaling coefficient equals the MAM scaling coefficient.

140

141 2.2. Scaling interpretation

142 2.2.1. Theoretical dissymmetry of scaling distribution

143 Whatever the scaling coefficient chosen ($b_{i,d}$, β_i or $\gamma_{i,d}$), its interpretation is not symmetric
 144 related to 1 (no scaling). Actually, when a scaling coefficient is at a distance of 0.9 from 1, it
 145 could be interpreted as:

- 146 - the panelist spreads 10 times less his scores than the panel ($1-0.9=0.1$).
- 147 - the panelist spreads his scores almost 2 times more than the panel ($1+0.9=1.9$).

148 This asymmetry can be problematic with computations involving averages or variances of
 149 scaling coefficients, like ANOVA. In this situation, we recommend transforming the scaling
 150 coefficients with a function f such as:

$$f: x \in [1; +\infty[\rightarrow x - 1$$

$$f: x \in]0; 1[\rightarrow 1 - \frac{1}{x}$$

$$f: x \in]-\infty; 0] \rightarrow NA$$

151 This function is such as the image of a scaling coefficient of 1 is 0, and such as

152 $f\left(\frac{1}{10}\right) = -9 = -f(10)$.

153 *[Insert Figure 3]*

154 This function is invertible with $f^{-1}: x \in]-\infty; 0] \rightarrow 1/(1 - x)$ and

155 $f^{-1}: x \in [1; +\infty[\rightarrow 1 + x$

156 2.2.2. Extreme values and scaling

157 Extreme values also impact averages and variances. Consequently, such extreme values
 158 should not be taken into account in most of the analyses of the scaling coefficients.
 159 Furthermore, too large or too small values of scaling are not interpretable as real scaling
 160 effects, but indicates a disagreement with the panel. Thus, a limit L was proposed to select
 161 only “reasonable values” for scaling coefficient (2 for overall scaling and 10 for corrected
 162 scaling, according to their related distributions, cf. section “Distribution of the different
 163 indicators”). Therefore, the transformation f was completed by a restriction on $[1/L; L]$,
 164 hence any scaling coefficient out of this interval was considered as a missing value.

165 In the rest of the paper, the resulted function f_L was systematically applied when variances
 166 and averages were calculated and compared.

167 2.3. Material for the meta-analysis

168 Balanced and complete datasets with continuous scales including at least 3 products were
169 extracted from the SensoBase (www.sensobase.fr) (Pineau, 2006) in order to calculate the
170 different scaling coefficients. Datasets with descriptors having a null variance or whose linear
171 regression involved in scaling calculation perfectly fits the data (*ie*, with null residuals) were
172 removed from the analysis for computational reasons.

173 Finally, 187 datasets were kept, representing 1 088 products, 1 676 panelists, 3 844
174 descriptors and 646 358 observations.

175 Age and gender effects on scaling were investigated based on the 154 datasets with
176 available information. Same investigation for the smoking status was conducted on 72
177 datasets.

178

179 2.4. Methods for the meta-analysis of scaling coefficients

180 2.4.1. Distribution of scaling coefficients

181 Individual and corrected scaling coefficients were calculated for the selected datasets. A
182 table containing percentages of negative, between 0 and $\frac{1}{2}$, $\frac{1}{2}$ and 1, 1 and 2, and higher
183 than 2 scaling coefficients was calculated.

184 The percentage of scaling coefficients between $\frac{1}{10}$ and 10 was also calculated for checking
185 appropriateness of an L-value of 10.

186 2.4.2. Analysis of overall scaling coefficient

187 For a given dataset, one-way ANOVA model of the panelist effect applied to $f_L(b_{i,d})$ was
188 used for testing whether the averages of the MAM scaling coefficients were similar across
189 panelists. The higher the percentage of datasets with a significant panelist effect, the more
190 the overall scaling coefficient is justifiable.

191 Then, one-way ANOVA was run for detecting the significance ($p=0.05$) of age, gender and
192 smoking status effects on overall scaling coefficient. When significant, the analysis was
193 completed by a multiple comparisons of means (HSD, $p=0.05$).

194 The stability within subject over datasets of overall scaling coefficients was studied for 41
195 panelists having at least 4 participations (involving a total of 33 datasets). The individual
196 standard deviations of their overall scaling coefficients were calculated (each one computed
197 on 4 or more values). If the panelist keeps the same overall scaling coefficient, the related
198 standard deviations should be close to 0. The standard deviation of the overall scaling
199 coefficients was also calculated between panelists. Thus, the between-panelist variability of
200 overall scaling and the within one could be compared. This aims to check whether overall
201 scaling coefficients are more similar within panelists than between panelists. Furthermore, a
202 potential panelist effect on overall scaling was tested with a one-way ANOVA. A significant

203 panelist effect would mean that some panelists are more or less heterogeneous in their
 204 overall scaling coefficients than the panel.

205 2.4.3. Analysis of corrected scaling coefficient according to sensory modality

206 Within a dataset, the standard deviation of the corrected scaling coefficients was calculated
 207 for each descriptor as a measure of individual heterogeneity of scaling. Then a one-way
 208 ANOVA of the sensory modality factor was run on these measures of heterogeneity. Multiple
 209 comparisons of means of the different sensory modalities were conducted by the HSD
 210 procedure ($p=0.05$). Therefore, potential strong heterogeneities in sensitivities for a given
 211 sensory modality could be detected.

212 3. Results: Meta-analysis of 187 profile studies

213

214 3.1. Distribution of the different scaling indicators

215 Table 1 shows the distribution of 52 881 MAM and corrected scaling coefficients and 2 381
 216 overall scaling coefficients. Of course, the number of overall scaling coefficients is much
 217 lower than the number of MAM and corrected scaling coefficients as they are one per
 218 panelist instead of one per panelist by descriptor.

219 [Include Table 1]

220 The most striking result is the drastic reduction of the percentage of negative overall scaling
 221 coefficients, compared to the percentage of negative MAM scaling coefficients.

222 Finally, 90% of the distribution of the overall scaling coefficients is between $\frac{1}{2}$ and 2,
 223 whereas 94% of non-negative specific and MAM scaling coefficient are between $\frac{1}{10}$ and
 224 10. This justifies our choice of L for the different scaling coefficients described in the section
 225 “Extreme values and scaling”.

226 3.2. Overall scaling and panelist characteristics

227 A significant panelist effect was found for MAM scaling coefficients in 72% of the datasets. It
 228 justifies our choice to consider an overall scaling coefficient unique by panelist.

229 Table 2 shows that over sixty years-old people spread their scores significantly less than the
 230 rest of the panel ($f_L(\beta)$ is negative).

231 [Include Table 2]

232 Table 3 shows that smokers spread significantly less their scores than the ex-smokers.
 233 Smokers spread also less their scores than the non-smokers, and the significance is almost
 234 reached ($p=0.069$). No significant differences are observed between ex-smokers and no-
 235 smokers ($p=0.54$).

236 [Include Table 3]

237 β can be explained as a psychological trend or physiological hypo or hyper-sensitivities for
 238 most of descriptors. Consequently, over sixty years old people and smokers are suspected to
 239 be less sensitive for most of the descriptors used in sensory studies, because of the loss of
 240 sensory abilities. These results validate the hypothesis of hyposensitivity of seniors and
 241 smokers presented in Introduction. Note that ex-smokers are in the same group than non-
 242 smokers, suggesting that they may have recovered most of their sensory abilities.

243 Regarding the study of stability of overall scaling over time, a panelist effect was observed
 244 for the overall scaling ($p < 0.0001$) suggesting that panelists do not have the same overall
 245 scaling coefficient. In average, the standard deviation of overall scaling coefficient by
 246 assessor was 0.30, whereas the standard deviation of overall scaling by dataset was 0.37 in
 247 average. Consequently, the standard deviation within panelist is smaller than the standard
 248 deviation between panelists, but not as much as expected.

249 3.3. Corrected scaling and sensory modalities

250 The p-value was significant (< 0.0001) for the sensory modality effect on the standard
 251 deviations of the corrected scaling coefficient. Table 4 shows more differences in sensitivity
 252 for Aroma, Taste and Smell, less in Texture and Visual.

253 [Insert Table 4]

254 It could reveal that the taste and odor sensations are more likely to be different from a
 255 panelist to another than somaesthetic or visual sensations. Therefore, the analysis of the
 256 data by sensory modality could sometimes be noteworthy.

257 4. Discussion

258 4.1. Interest of scaling decomposition

259 The strength of MAM model is to systematically improve the discrimination of the panel.
 260 However, if a panelist really feels higher differences between the products than the panel,
 261 (*ie* the differences are not due to a bias of use of scale) the MAM over- corrects the scores.
 262 Using the overall scaling coefficient allows correcting the use of the scale for all descriptors
 263 at the same time and reduces the risks of such an artificial improvement of discrimination.

264 4.2. Calculation of overall scaling coefficient

265 We might have considered the average MAM scaling coefficients instead of the regression of
 266 all confounded scores. However, the average of MAM scaling coefficients would have been
 267 very sensitive to extreme values, and particularly to negative MAM scaling coefficients.
 268 Calculating the average of the MAM coefficients gives the same weight to each descriptor,
 269 even if some of them do not discriminate the products. Yet, negative MAM scaling
 270 coefficients occur especially when products are not discriminated by the descriptors.

271 Consequently, the average of MAM scaling coefficients could be very sensitive to a non-
 272 discriminative descriptor. In this case, the average would be smaller than it should be.
 273 Overall scaling coefficient solves this problem as the data point from non-discriminative
 274 descriptors would be central in the scatterplot of the regression. On the contrary, very
 275 discriminative descriptors would have a great importance in the regression. Finally, the
 276 sensibility of the regression to extreme points is more an advantage than an inconvenient.

277 One could argue that scaling should not be linear. Although our regression is linear, other
 278 models could be chosen to obtain information regarding the use of the scale. Thus, further
 279 work could compare different models for approximating the panelist scores with the panel
 280 scores.

281 4.3. Transformation of scaling coefficient

282 Another point of discussion could be the transformation of scaling required to analyze
 283 variance of scaling coefficients. A logarithmic function (\ln) would have respected the
 284 conditions ($\ln(1) = 0$ and $\ln(1/x) = -\ln(x)$). However, our choice of function allows
 285 keeping an intuitive idea of our results. For example, a panelist spreading his scores 2 times
 286 more (or less) than the panel would have a scaling coefficient of 2 (or $\frac{1}{2}$). The image of 2
 287 under f_L (or $\frac{1}{2}$) gives 1 (or -1), whereas the logarithm function would give $\ln(2) \approx 0.69$ (or -
 288 0.69) more difficult to interpret. On the contrary, an image of 3 (or -3) means that the
 289 panelist spreads 4 times more/or less than the panel.

290 4.4. Dealing with negative overall scaling coefficients

291 Negative overall scaling coefficients rarely occur but are still present. If neither the panelist
 292 nor the panel discriminates the products for all descriptors, the negativity of one coefficient
 293 is understandable: on a continuous scale, giving exactly the same score to identical products
 294 is difficult. The given scores can also be reversed compared to those of the panel and this
 295 situation results in a negativity issue.

296 If the products were discriminated by the panel, a discussion with the panelist regarding the
 297 scoring task should take place to determine whether the panelist understood the labels of
 298 the scale for all descriptors (to check if he reversed his scores). The panelist could also have a
 299 sensory problem for most of the descriptors of the study, or having not seriously scored the
 300 products.

301 4.5. Stability of overall scaling over time

302 The variability of overall scaling over time was rather unexpected. Indeed, if this coefficient
 303 reflects a trend for all descriptors, it should be stable over time. Several reasons could
 304 explain overall scaling variations over time. Firstly, the calculation of overall scaling
 305 coefficient depends on the panel scores. Consequently, the scaling coefficient of one
 306 panelist can change across datasets only because the panels involved in these datasets are

307 different. Secondly, instructions and scale labels vary with datasets, inducting a possible
 308 change in the panelist behavior. Finally, the mood of panelists varies depending on time and
 309 could impact the scaling effect.

310 4.6. Consequences of differences of variability between sensory modalities

311 As Table 4 showed that sensory modality impacts the scaling effect, overall scaling could also
 312 be calculated by sensory modality. Thus, it would be possible to see whether a panelist
 313 would be more sensitive to visual, texture, odor or aroma descriptors. Several profiles of
 314 panelist sensitivities could exist, and this information would certainly be useful to the panel
 315 leader while recruiting his panel for profiling a specific type of products involving some
 316 sensory modalities more than others.

317 Another consequence of the differences of variability between sensory modalities is that, if a
 318 subject is less sensitive to a specific sensory modality than others, the scaling coefficients
 319 could depend on the constitution of the descriptor set. For instance, let's imagine a panelist
 320 hyposensitive to texture only: the more texture attributes in the descriptor set, the lower his
 321 overall scaling coefficient (and the higher his corrected scaling coefficients).

322 To overcome this potential problem, the analysis can be derived by sensory modality,
 323 computing both types of scaling coefficients only within modalities. Thus, the overall scaling
 324 coefficient for a given sensory modality indicates a global trend of hypo or hyper sensitivity
 325 for this specific sensory modality and the corrected scaling coefficient takes that into
 326 account. This approach protects against over or under representation of a sensory modality
 327 in the global list of attributes. Of course, this information of sensitivity relative to each
 328 sensory modality is of interest for monitoring panel performances.

329 4.7. Opening to other performances

330 Decomposing a panelist performance in two components, a total one (regardless of
 331 descriptors) and a descriptor-specific one, could be adapted to the panelist effect (to
 332 measure differences among panelists in their mean scores). Thus, the component common
 333 to all descriptors would represent a trend to systematically give high (or low) scores
 334 (regardless of descriptors) and the component specific to one descriptor would indicate the
 335 absolute sensitivity to this descriptor (in opposition to the relative sensitivity detected in
 336 scaling effect).

337 4.8. Multivariate extension of the Mixed Assessor model

338 Unlike Romano et al. (2015) with the multiplicative Assessor Model, this paper does not
 339 propose a multivariate extension of the Mixed Assessor Model, but uses several regression
 340 coefficients in order to give precious indications about scaling (physiological and
 341 psychological decomposition). Further work could integrate the overall scaling coefficient
 342 into a multivariate statistical model.

343 5. Conclusion

344 The MAM scaling coefficient can be decomposed into an overall scaling coefficient
 345 (descriptor independent) and a corrected scaling coefficient (descriptor-specific). Overall
 346 scaling is more likely to represent the psychological scaling effect whereas corrected scaling
 347 aims to represent physiological effect specific to a given descriptor (hyper- or hypo-
 348 sensitivity relatively to the panel). Practically, meta-analyses showed that people over sixty
 349 years old and smokers tend to spread less their scores than other panelists. Furthermore,
 350 the use of overall scaling solves the problem of negativity of the scaling coefficients.
 351 Consequently, using the scaling decomposition is recommended in order to find out causes
 352 of scaling and potentially correct them.

353 As a perspective, statistical tests should be set up to test significance of overall scaling. In
 354 this purpose, the overall scaling coefficient will be included in a multidimensional extension
 355 of the Mixed Assessor Model.

356 **Acknowledgements**

357 We are grateful to the Region of Burgundy which funded our researches with the JCE grant.

358 Ahlström, R., Berglund, B., Berglund, U., Engen, T., & Lindvall, T. (1987). A comparison of odor
 359 perception in smokers, nonsmokers, and passive smokers. *American Journal of*
 360 *Otolaryngology*, 8(1), 1-6.

361 Booth, P., Kohrs, M. B., & Kamath, S. (1982). Taste acuity and aging: A review. *Nutrition Research*,
 362 2(1), 95-109.

363 Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account
 364 by analyzing profile data with the Mixed Assessor Model. *Food Quality and Preference*, 39(0),
 365 156-166.

366 Brockhoff, P. M., & Skovgaard, I. M. (1994). Modelling individual differences between assessors in
 367 sensory evaluations. *Food Quality and Preference*, 5(3), 215-224.

368 Herbert, S., & L., S. J. (2004). *Sensory evaluation practices (Third edition)*. San Diego, California (USA):
 369 Elsevier Academic Press.

370 Krut, L. H. P., M.; J., Bronte-Stewart, M.D.,. (1961). Taste perception in smokers and non-smokers.
 371 *British Medical Journal*, 1:384.

372 Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference*, 22(8),
 373 733-747.

374 Meilgaard, M., Gail Vance Civille, B. S., & Carr, T. B. (1999). *Sensory evaluation techniques (3. ed.)*.
 375 USA: Taylor & Francis Group LLC.

376 Peltier, C., Brockhoff, P. B., Visalli, M., & Schlich, P. (2014). The MAM-CAP table: A new tool for
 377 monitoring panel performances. *Food Quality and Preference*, 32, Part A(0), 24-27.

378 Pineau, N. (2006). *La performance en analyse sensorielle, une approche base de données*.

379 Romano, R., Næs, T., Brockhoff, P. B., Romano, R., Næs, T., & Brockhoff, P. B. (2015). Combining
 380 analysis of variance and three-way factor analysis methods for studying additive and
 381 multiplicative effects in sensory panel data. *Journal of Chemometrics*, 29(1), 29-37.

382

383

Table 1: Distribution of the different scaling coefficients

	n	$\beta < 0$	$0 \leq \beta < 0.5$	$0.5 \leq \beta < 1$	$1 \leq \beta < 2$	$\beta > 2$
MAM β	52 881	16.12%	18.16%	21.37%	29.74%	14.61%
Overall β	2 381	0.63%	7.98%	41.83%	48.09%	1.47%
Corrected β	52 881	16.12%	16.08%	22.33%	31.30%	14.17%

Table 2: Effect of age on overall scaling

Age	n	Mean $f_L(\beta)^1$	Related β^2
18-30	214	0.047 ^a	1.047
30-40	211	0.016 ^a	1.016
40-50	409	0.041 ^a	1.041
50-60	392	0.027 ^a	1.027
60-70	380	-0.071 ^b	0.934
+70	120	-0.075 ^b	0.930

¹: Two means with the same letter are not significantly (LSD, p=0.05) different

²: Related β is $F^{-1}(\text{Mean } f_L(\beta))$ and can be understood as a usual beta

Table 3: Effect of smoking status on overall scaling

	n	Mean $f_L(\beta)^1$	Related β
Non smoker	590	0.001 ^{ab}	1.001
Ex smoker	149	0.033 ^a	1.033
Smoker	68	-0.096 ^b	0.912

¹: Two means with the same letter are not significantly (HSD, p=0.05) different

²: Related β is $f_L^{-1}(\text{Mean } f_L(\beta))$ and can be understood as a usual beta

Table 4: Effect of sensory modality on standard deviation of corrected scaling

Sensory modality	Number of descriptors	Mean $Sd(f_L(\beta))^1$
Appearance	335	1.22 ^a
Texture	703	1.24 ^a
Trigeminal	107	1.63 ^b
Basic taste	741	1.67 ^b
Flavor (retronasal)	1150	1.81 ^{bc}
Aftertaste	129	1.85 ^{bc}
Odor (orthonasal)	302	1.89 ^c

¹: Two means with the same letter are not significantly (LSD, $p=0.05$) different

Figure 1: Obtaining the MAM scaling coefficient corresponding to a given panelist i and to a given descriptor p

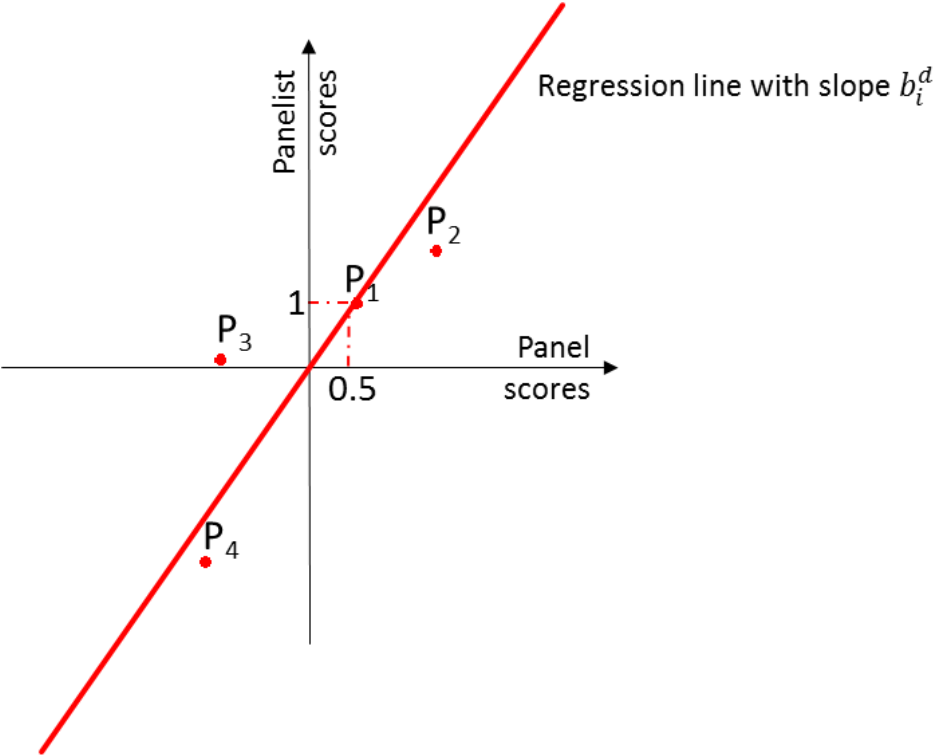
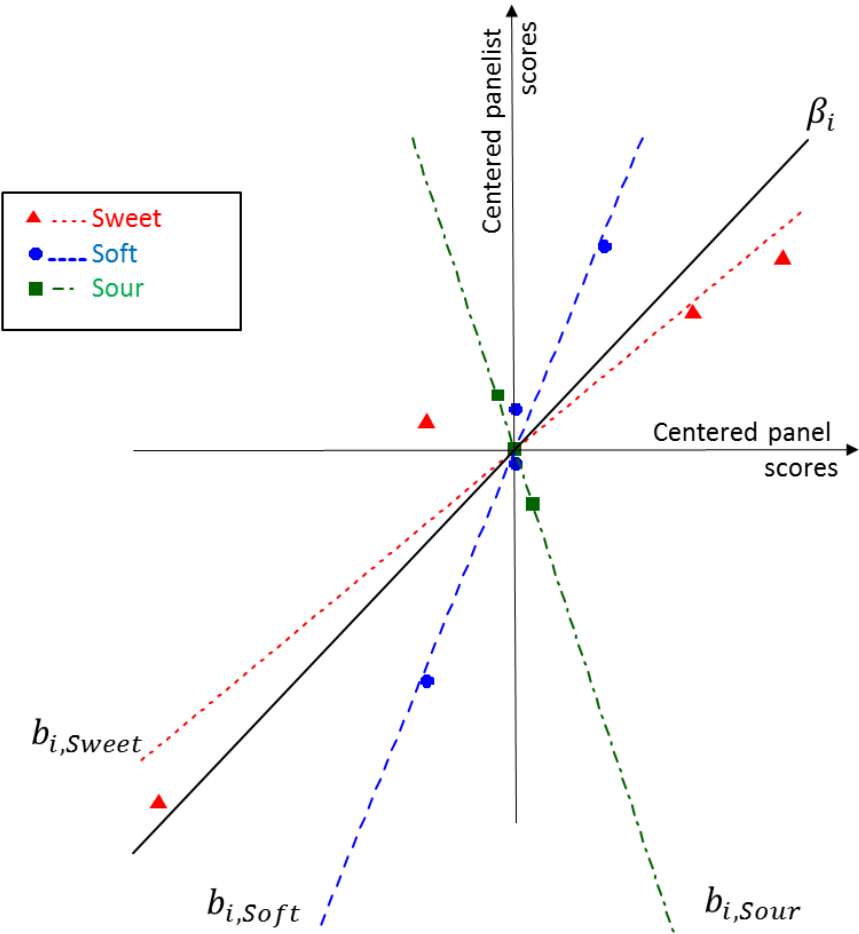
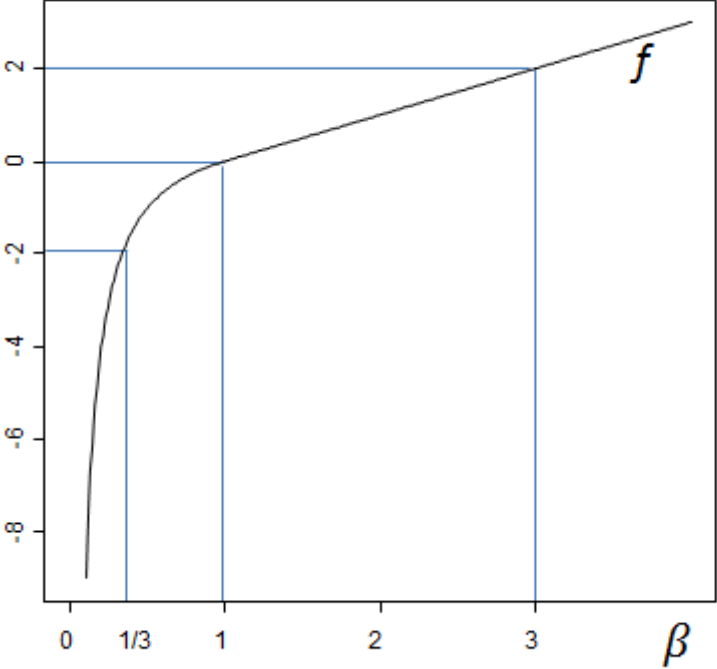


Figure 2: Obtaining the overall scaling coefficient β_i



$b_{i,Sweet}$, $b_{i,Soft}$ and $b_{i,Sour}$ are the slopes of the regressions by descriptor (MAM scaling coefficients).
 β_i is the slope of the regression all descriptors being pooled

Figure 3: The f transformation for scaling coefficients





Dealing with individual differences in the use of the scale: the MultMAM R-package

Caroline Peltier
CNRS¹, INRA², UB³, E2S⁴

Michel Visalli
CNRS¹, INRA², UB³

Pascal Schlich
CSGA¹, INRA², UB³

¹CNRS, UMR 6265 CSGA, Centre des Sciences du Goût et de l'Alimentation, Dijon, France

² INRA UMR 1324 Centre des Sciences du Goût et de l'Alimentation , Dijon, France

³ UB Université de Bourgogne Franche-Comté, UMR CSGA, Dijon, France

⁴E2S, Dijon, France

Abstract

Sensory profiling aims to describe the sensory characteristics of food products using a list of descriptors with a panel of trained assessors. During sensory evaluations (or any other scoring task), individual differences in the scale width effectively used by assessors (scaling effect) are regularly observed. The Mixed Assessor Model (MAM) proposed by Brockhoff, Schlich, and Skovgaard (2015) takes the scaling effect into account in a univariate statistical model including a scaling coefficient per assessor.

Peltier, Visalli, and Schlich (2015b) proposed a multivariate extension to the scaling coefficient:

- (i) with different scaling coefficients varying with descriptors,
- (ii) with a single coefficient of scaling calculated by assessor and descriptor. In the same paper, these scaling coefficients were studied on 187 datasets, but were not included in a multivariate statistical model.

The current paper presents a multivariate model (the MultiMAM) extending the MAM and including both approaches of scaling coefficient (overall or not).

This paper offers the package MultMAM containing functions to obtain the different scaling coefficients and to build maps taking scaling effect into account with MultiMAM (MAM-CVA). The use of the package is illustrated with an example based on a dataset of sensory profiling.

We recommend using overall MAM-CVA to obtain a product map in sensory profiling. This model (and consequently this package) can also be used with any type of data where human beings score several characteristics on a scale (teaching, marketing, psychology...)

Keywords: MAM, multivariate, maps, scaling, CVA.

1. Introduction

In several disciplines, the human being has to score items on continuous scales. For instance, sensory profiling aims to describe the sensory characteristics of food products with a panel using a list of descriptors. Assessors involved in a sensory profiling task are trained to rate the perceived intensities of a number of descriptors such as: sweetness, saltiness, crunchiness etc. The obtained data is usually analyzed by ANalysis Of VAriance (ANOVA) - to detect differences between products - and Principal Component Analysis (PCA)- to map the product space-. However, in these analyses, the fact that an assessor spreads more (or less) his scores than the panel -the so-called **scaling effect** - is not taken into account.

The scaling effect was first presented in Brockhoff and Skovgaard (1994) with the Assessor Model. However, this statistical model did not provide a test for the product effect and was consequently not used in sensory analysis. Recently, Brockhoff *et al.* (2015) proposed a model taking the scaling effect into account allowing the product effect to be tested: the Mixed Assessor Model (MAM). This model was used by Peltier, Brockhoff, Visalli, and Schlich (2014) to produce the MAM-CAP table, a visual tool for monitoring panel and panelist performances for each descriptor.

A decomposition of the MAM Scaling effect taking all descriptors into account was also proposed by Peltier *et al.* (2015b). It differentiates two scaling components:

- The **overall scaling**, independent of the descriptors: this scaling coefficient is calculated by assessor, all descriptors confounded. It could be interpreted as a "*psychological part*" of scaling effect.
- The **corrected scaling**, dependent of the descriptors: it could assess a specific sensitivity of the assessor for some descriptors and be interpreted as a "*physiological part*" of scaling effect.

In this paper, a multidimensional model related to the univariate MAM is presented. This model considers either the MAM scaling (without considering the different causes of scaling) or the "overall scaling" (correcting a potential scaling effect only if it is common to all descriptors). This multivariate model allows maps taking properly the scaling (overall or not) effect into account to be plotted. The R functions related to the multidimensional extension of the MAM are available in the **MultMAM** package. The **MultMAM** functions are described and illustrated with a sensory profiling example, but the scaling effect is not specific to sensory profiling and this package could also be applied for any other field implying panelists scoring different items on a scale (teaching, marketing, psychology...).

2. Methodological background

2.1. Theory of univariate MAM

Notations and classical model in sensory profiling

Highlighting differences between products is the objective of any sensory analyst performing a sensory profiling. Attesting for significant differences between products requires a model.

The classical model used in sensory analysis is the two-way product*assessor model with random assessor effect (Naes 1990). Therefore, for each descriptor d , assessor i , product j , and replicate k , the score y_{ijk}^d can be decomposed into the following components:

$$y_{ijk}^d = \mu^d + \alpha_i^d + \nu_j^d + c_{ij}^d + \varepsilon_{ijk}^d \quad (1)$$

where μ^d is the average for the descriptor d , α_i^d the assessor effect, ν_j^d the product effect, c_{ij}^d the interaction term and ε_{ijk}^d the error term: it is the mixed model. We note y^d , $y_{i..}^d$, $y_{.j.}^d$ and $y_{ij.}^d$ the observed averages (overall, by assessor, by product and by product and assessor). As a consequence of the mixed model, the product effect is tested against the interaction effect to determine statistical differences between products (**discrimination**).

The scaling effect according to Brockhoff

Calculating the standard deviations of each assessor could be the first idea to find out whether an assessor spreads his scores more than the others or not. However, this approach consists of calculating $\sqrt{\frac{1}{JK-1} \sum_{j,k} (y_{ijk}^d - y_{i..}^d)^2}$. Such a quantity can correspond to:

- (i) differences between products
- (ii) differences between replicates for the assessor
- (iii) a trend to spread his scores (scaling effect)
- (iv) a combination of (i) (ii) and (iii).

Consequently, computing standard deviations is not specific enough and can not be the right approach to determine a scaling effect. To remove the differences due to (ii), a solution could be to calculate $\sqrt{\frac{1}{J-1} \sum_j (y_{ij.}^d - y_{i..}^d)^2}$ instead of the classical standard deviations. But, this quantity still contains variability due to (i) and (iii).

Brockhoff *et al.* (2015) proposes to regress the centered scores for each assessor ($y_{ij.}^d - y_{i..}^d$) on the centered panel scores ($y_{.j.}^d - y^d$). If the slope of this regression \tilde{b}_i^d equals 1, there is no scaling effect. If $\tilde{b}_i^d > 1$ (or $0 < \tilde{b}_i^d < 1$), the assessor i spreads more (or less) his scores than the panel for the descriptor d . This coefficient is a relevant indicator of the scaling effect.

If $\tilde{b}_i^d < 0$, it means that the assessor scored the products inversely of the panel. It can happen for 3 reasons:

- when the assessor totally disagrees with the panel,
- when the products are not discriminated enough (in this case the product order given by the panel is random and can be the opposite of the product order given by the panelist),
- when the assessor totally misunderstood the scale.

The Mixed Assessor Model

Brockhoff *et al.* (2015) includes the scaling effect in the Mixed Assessor Model (MAM), considering the following decomposition of scores:

$$y_{ijk}^d = \mu^d + \alpha_i^d + \nu_j^d + b_i^d x_j^d + \delta_{ij}^d + \varepsilon_{ijk}^d \quad (2)$$

where μ^d is the average for the descriptor d , α_i^d the assessor effect, ν_j^d the product effect $x_j^d = y_{.j}^d - y_{...}^d$ the centered panel scores, δ_{ij}^d the "pure disagreement", and ε_{ijk}^d the error term.

b_i^d is the slope of the linear regression of the interaction on the centered panel scores.

Thus, b_i^d is such as the quantity

$$\|(y_{ij.}^d - y_{i..}^d - y_{.j}^d + y_{...}^d) - b_i^d(y_{.j}^d - y_{...}^d)\| \quad (3)$$

is minimized, where $\|\cdot\|$ is the euclidean norm.

$$b_i^d = \frac{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)(y_{ij.}^d - y_{i..}^d - y_{.j}^d + y_{...}^d)}{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)^2} \quad (4)$$

Furthermore, b_i^d can also be written:

$$b_i^d = \frac{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)((y_{ij.}^d - y_{i..}^d) + (-y_{.j}^d + y_{...}^d))}{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)^2} = \frac{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)(y_{ij.}^d - y_{i..}^d) - (y_{.j}^d - y_{...}^d)^2}{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)^2} \quad (5)$$

$$b_i^d = \frac{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)(y_{ij.}^d - y_{i..}^d)}{\sum_{j=1}^J (y_{.j}^d - y_{...}^d)^2} - 1 = \tilde{b}_i^d - 1 \quad (6)$$

where \tilde{b}_i^d is the coefficient of the regression of centered assessor scores on the centered panel scores presented in 2.1.2. Thus, the difference between MAM and classical two-way model is the decomposition of the classical interaction between variability due to a different use of the scale (scaling effect represented by $b_i^d x_j^d$) and the "pure disagreement" represented by δ_{ij}^d .

$$y_{ij.}^d - y_{.j}^d - y_{i..}^d + y_{...}^d = b_i^d * x_j^d + \delta_{ij}^d. \quad (7)$$

Therefore, with the unidimensional MAM, a score is a combination of a general mean, a product effect, a subject effect, a scaling effect, a pure disagreement effect and an error term.

Note that this model requires at least 3 products (to calculate a regression based on at least 3 points) to avoid perfect fitting. 2 evaluations are required to have an interaction term in the model, but the calculation of scaling coefficient is also possible with dataset with only one replicate.

Furthermore, Brockhoff *et al.* (2015) recommend not to correct the scaling effect and to consider the classical model when $\tilde{b}_i^d (= b_i^d + 1)$ is negative.

2.2. Multivariate MAM

A matrix formulation of the MAM

The principle of the multivariate MAM (MultiMAM) is to use vectors instead of scalars in the univariate model. Let's consider the vector of the scores of all descriptors

$$Y_{ijk} = \begin{pmatrix} y_{ijk}^1 \\ y_{ijk}^2 \\ \dots \\ y_{ijk}^D \end{pmatrix}$$

The natural extension of the unidimensional MAM could be written:

$$Y_{ijk} = \mu + \alpha_i + \nu_j + \beta_i X_j + \delta_{ij} + \varepsilon_{ijk}$$

where $\mu = \begin{pmatrix} \mu^1 \\ \mu^2 \\ \dots \\ \mu^D \end{pmatrix}$: vector of the averages; $\alpha_i = \begin{pmatrix} \alpha_i^1 \\ \alpha_i^2 \\ \dots \\ \alpha_i^D \end{pmatrix}$: panelist effect vector;

$\nu_j = \begin{pmatrix} \nu_j^1 \\ \nu_j^2 \\ \dots \\ \nu_j^D \end{pmatrix}$: product effect vector; $\delta_{ij} = \begin{pmatrix} \delta_{ij}^1 \\ \delta_{ij}^2 \\ \dots \\ \delta_{ij}^D \end{pmatrix}$: pure disagreement ;

$\varepsilon_{ijk} = \begin{pmatrix} \varepsilon_{ijk}^1 \\ \varepsilon_{ijk}^2 \\ \dots \\ \varepsilon_{ijk}^D \end{pmatrix}$: error term; $X_j = \begin{pmatrix} y_{.j}^1 - y_{...}^1 \\ y_{.j}^2 - y_{...}^2 \\ \dots \\ y_{.j}^D - y_{...}^D \end{pmatrix}$ and β_i to be determined below.

First option: the MAM scaling coefficient

A first idea is to consider β_i as the vector constituted of the univariate b_i^d .

$$\beta_i = \begin{pmatrix} b_i^1 \\ b_i^2 \\ \dots \\ b_i^D \end{pmatrix}$$

This model is called "MultiMAM". In this case, the product between β_i and X_j corresponds to the R product between two vectors (one to one). This model considers the scaling as depending on the descriptor, and corresponds strictly to the natural extension of the univariate Mixed Assessor Model.

Second option: the "overall" scaling coefficient

A second idea would be to consider β_i as unique by subject (overall scaling).

In this purpose, the univariate methodology is extended to multivariate one. The multivariate extension of (3) requires a norm of matrices. We choose the Froebenius norm $\|\cdot\|_F$ (such

as $\|M\|_F = \sqrt{\sum_{i,j} m_{ij}^2}$, also used to extend the Pearson correlation coefficient to the RV coefficient (Escoufier 1973).

Let's consider β_i minimizing $\|(Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...}) - \beta_i(Y_{.j.} - Y_{...})\|_F$. This β_i allows to correct scaling only when it is common to all descriptors. Note that the obtained $\tilde{\beta}_i = \beta_i + 1$ equals the "overall scaling" calculated in Peltier *et al.* (2015b).

$\tilde{\beta}_i$ corresponds to the calculation of the coefficient of the regression of the assessor scores (centered by descriptor and all descriptors confounded) on the centered panel scores. Consequently, $\tilde{\beta}_i$ illustrates a global trend of the panelist to spread his scores whatever the descriptor.

In this case, the product of β_i and X_j corresponds to the product * of R between one scalar and one vector.

Finally, two options exist in the multivariate extension of the MAM:

- the first one considers the classical MAM scaling, depending on descriptors and corresponds to the "mam" option,
- the second one considers the individual scaling depending on panelist only, and corresponds to the "overall" option.

In sensory analysis, we recommend to use the second one. Indeed, the MultiMAM corrects the scaling for each descriptor separately without taking the whole data into account : both psychological and physiological components are corrected. The Overall MAM aims to correct only the psychological component of scaling and considers the physiological variations as disagreement between assessors. This approach of overall scaling can be extended to other fields where several descriptors are scored and be useful to study the behavior of assessors in the scoring tasks.

Decomposition of multivariate scores

Whatever the chosen option for β_i (overall or not), a vector of scores can be decomposed into the sum of a vector of means, a vector of product effect, a vector of subject effect, a vector of scaling effect, a vector of pure disagreement effect and a vector of residual such as following (estimation of the parameters of the model):

$$Y_{ijk} = Y_{...} + (Y_{i..} - Y_{...}) + (Y_{.j.} - Y_{...}) + \beta_i * (Y_{.j.} - Y_{...}) + D_{ij} + (Y_{ijk} - Y_{ij.})$$

$$\text{where } Y_{...} = \begin{pmatrix} y_{...}^1 \\ y_{...}^2 \\ \dots \\ y_{...}^D \end{pmatrix}, Y_{i..} = \begin{pmatrix} y_{i..}^1 \\ y_{i..}^2 \\ \dots \\ y_{i..}^D \end{pmatrix}, Y_{.j.} = \begin{pmatrix} y_{.j.}^1 \\ y_{.j.}^2 \\ \dots \\ y_{.j.}^D \end{pmatrix}, Y_{ij.} = \begin{pmatrix} y_{ij.}^1 \\ y_{ij.}^2 \\ \dots \\ y_{ij.}^D \end{pmatrix},$$

and $D_{ij} = (Y_{ij.} - Y_{.j.} - Y_{i..} + Y_{...}) - \beta_i * (Y_{.j.} - Y_{...})$

$\beta_i * (Y_{.j.} - Y_{...})$ is called "the scaling term" and D_{ij} the "pure disagreement term".

The (I*J*K, D) matrix (with IJK lines and D columns) Y containing the scores Y_{ijk} (with observations as lines and descriptors as columns) can be decomposed into the sum of (I*J*K, D) matrices corresponding to the estimations of the different effects.

$$Y = M + P + A + S + D + E \quad (8)$$

where M , P , A , S , D and E are $(I \times J \times K, D)$ matrices corresponding respectively to the mean, product, assessor, scaling, disagreement and error term.

2.3. Plotting a map in accordance with the MAM

Building a map allows a multidimensional space to be summarized in a visual way. In sensory analysis, such a map allows products, attributes, and even assessors to be represented on the same graph as a synthesis of the whole experiment.

Principle of the CVA

In sensory profiling, Principal Component Analysis (PCA) is generally used to plot product maps. PCA functions are available in many R packages like PCA in **FactoMineR** or princomp in **stats**. PCA is based on the table of product means, and aims to represent the map with the more **variability** between the product means. Canonical Variate Analysis (CVA) aims to represent the map with the more **multivariate discrimination** between the products.

Regarding the difference between discrimination and variability, in the two-way (product and assessor) univariate model 1, the F statistic of the product effect is the ratio between the variance due to the product effect and the variance due to the interaction product*assessor effect. It allows **unidimensional discrimination** between products to be detected. The natural extension of the unidimensional variance of product (respectively interaction) effect is the covariance matrix of product (or interaction) effect, noted B (or W_{int}). Consequently, the natural extension of the unidimensional F is the discrimination matrix $W_{int}^{-1}B$. This "discrimination matrix" is used in MANOVA to test the multivariate discrimination.

To build a CVA, the discrimination matrix is diagonalized and the obtained eigenvectors determine the axes to be plotted. The products are projected on the chosen axes and the attribute positions are determined. For more details, refer to [Peltier, Visalli, and Schlich \(2015a\)](#).

To extend CVA with Multi-MAM: the MAM-CVA

The MAM improves the discrimination of the products by removing the scaling term from the interaction term, the remaining term being called "pure disagreement term". Extending the CVA with the MultiMAM model consists in using the pure disagreement term (instead of the interaction term) in the multivariate discrimination.

In this purpose, the decomposition (8) is used. $B = P'P$ (covariance matrix of product effect) and $W = D'D$ (covariance matrix of disagreement effect) are calculated. $W^{-1}B$ is the discrimination matrix taking the scaling effect into account. Then, as in usual CVA, this matrix is diagonalized in order to produce eigen vectors and values. Eigen vectors give the axes of the map. Product coordinates in this map are obtained by projecting the product scores on the eigen vectors. As in the usual CVA, a biplot or a graphic with two maps can be plotted ([Peltier et al. 2015a](#)). This mapping method is named "MAM-CVA".

The obtained map represents the best discrimination taking the scaling effect (overall or not) into account. The MAM-CVA represents the best discrimination between products adjusted to scaling effect by descriptor. The overall MAM-CVA represents the best discrimination between products adjusted to overall scaling effect.

Note that the inversion of the W matrix could be very sensitive to colinearities in the matrix.

If the matrix is not well conditioned (condition number > 30 according to [Peltier et al. \(2015a\)](#)), the CVA should be considered carefully.

Statistical tests and MAM-CVA

A Multivariate ANalysis Of VAriance (MANOVA) is conducted to obtain the significance of the product effect at a multivariate level. However, specifying the MultiMAM model was not possible in the R `manova` function (which takes into account only fixed effects), and we consequently had to reprogram this function to adapt it to the MultiMAM model. According to the SAS procedure GLM, MANOVA only needs three parameters to calculate an F and a pvalue. These parameters are:

- M_1 , the covariance matrix of the effect to be tested
- M_2 , the covariance matrix of the effect to be compared with
- df_{M_2} , the degrees of freedom related to M_2

Note that the degrees of freedom related to M_1 are not required for the calculation of the pvalue of the MANOVA. Consequently, we programmed a function taking only M_1, M_2 and df_{M_2} as parameters and returning the same results as `manova` for the classical models (one way, two way, two way with random effect). We used this function to calculate a F appropriate to the MultiMAM by considering ($M_1 = B$, $M_2 = W$ and $df_{M_2} = (I - 1)(J - 2)$). Note that the calculation of the $(J - 1)$ different scaling coefficients led to a loss of $(J - 1)$ degrees of freedom. Consequently, as the interaction term had initially $(I - 1)(J - 1)$, the disagreement term would have $(I - 1)(J - 1) - (I - 1) = (I - 1)(J - 2)$ degrees of freedom.

Confidence ellipses in accordance with the model

Ellipses are used to graphically determine multidimensional differences between products : intersected ellipses indicates similar products.

Classically, confidence ellipses ([Saporta 1986](#)) around the product means represent the dispersion of the assessors. In this purpose, the assessor scores averaged on evaluations (and consequently corresponding to the matrix $P + A + S + D$) are directly projected on the map (IJ points) and a confidence zone is plotted.

However, this approach is not in adequation with the unidimensional approach : to establish differences between products, we compare their variability to the interaction product*subject (corresponding to the disagreement between subjects). In our package, the ellipses around the product means does not represent the dispersion of the assessors, but the disagreement of the assessors for this product. The scaling effect, but also the assessor effect are removed from the variability. Thus, the assessor scores without the assessor effect and the scaling effect (corresponding to the matrix $P + D$) are projected on the map. Ellipses are consequently smaller with our method, and this approach is in agreement with the chosen multivariate model.

Note that this remark is also valuable for the two-way model (without taking scaling effect into account). Then, removing the assessor effect from the data amounts to center the scores by assessor and descriptor. Once done, the projected resulted scores conduct to smaller ellipses.

The dimensionality test

The test of dimensionality of the product space in the CVA based on the likelihood ratio proposed by [Mardia \(1989\)](#) can also be adapted to the MultiMAM.

$$df_W(\lambda_{r+1}^2 + \dots + \lambda_D^2) \sim \chi_u^2$$

where $u = pk - p(r + 1) - (k - r - 1)r = (p - r)(k - r - 1)$; k is the number of products, p the number of descriptors; r the number of dimensions of the products space under H_0 : "The dimension of the product space is lower or equal to r ."; df_W is the number of degrees of freedom related to W in the model and λ_i the eigen values of $W^{-1}B$.

Hotelling table related to this map

The Hotelling test is the multivariate extension of the Student t-test. It compares two matrices (M_1 and M_2) in order to attest significant differences between them. This test calculates the inverse of $M = M_1' M_1 + M_2' M_2$. Hotelling test is present in several packages, but we use our own function in order to deal with no invertible matrices. If M is not invertible and $M_1 = M_2$, then the returned p-value is 1. On the contrary, if one mean differs from the other, the return p-value is 0. In both cases, the returned statistic is NA.

Thus, to compare two products P_1 and P_2 , the matrices containing only the scores related to P_1 noted $X_{|P_1}$ (or related to P_2 noted $X_{|P_2}$) are used. [Peltier et al. \(2015a\)](#) recommend to conduct this test only on the significant dimensions of the product space. Indeed, if Hotelling T2 test was performed on all axes, too much importance might be given to the noise in the data: within variability will increase and so products could wrongly said to be equal. To compare two products according to MultiMAM, we would like to compare the variability due to the product effect and the variability due to the pure disagreement. For this purpose, the used matrices are $(P + D)_{|P_1}$ and $(P + D)_{|P_2}$.

A table containing the p-value of Hotelling test for all products pairs is drawn: the Hotelling table. When a p-value is lower than $\alpha = 0.05$, the two related products are significantly different.

When two products are not significantly different, a segment link them on the CVA (or MAM-CVA) map.

3. Examples with a cheese dataset

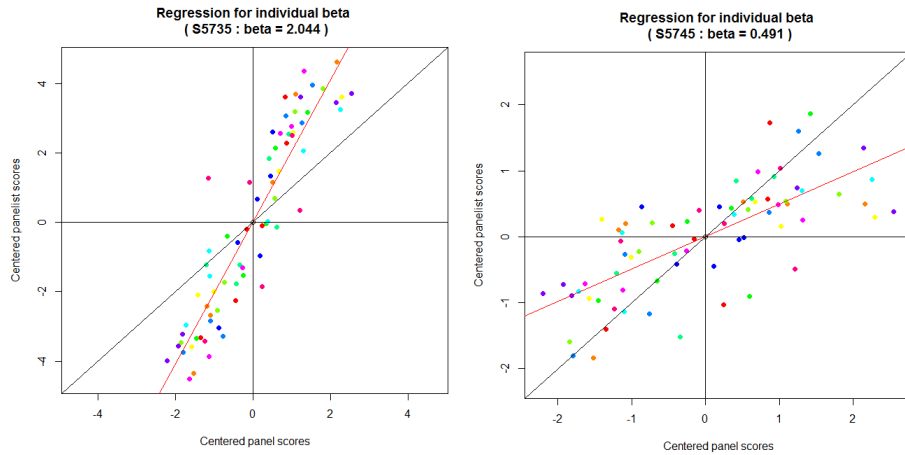
3.1. Preliminary

In this section, the **MultMAM** functions were applied to an example dataset. In this purpose, the **MultMAM** package and the example data `cheeses` were loaded with the following commands:

```
load("MultMAM"); data(cheeses);
```

The dataset `cheeses` was extracted from the `SensoBase`, a sensory database containing more than 1000 profile datasets. Six cheeses were scored by 16 subjects over a continuous scale (from 0 to 10) on 12 descriptors, in triplicate.

Figure 1: Example of large and small overall scaling coefficient



3.2. Analyzing the dataset with MultiMAM function

MultiMAM function was applied to the dataset (`resMAM=MultiMAM(cheeses)`). Three options can be chosen for the `modelType` ("overall", "mam" or "tw"). By default, the `modelType` is "overall", and the MultiMAM function returns a list containing several outputs.

```
>names(resMAM) [1] "decomposition" "beta" "avgMat" "trtMat" "assMat"
[6] "scalMat" "disagMat" "errMat" "intMat" "nNeg" "pvalScal"
```

`resMAM$Beta` gives the scaling coefficient as following:

S5733	S5734	S5735	S5736	S5737	S5738	S5739
1.5071705	0.9472210	2.0437441	1.0040833	1.4438035	0.9363514	0.7812305
S5740	S5741	S5742	S5743	S5744	S5745	S5746
0.6536565	0.5997861	1.1245597	0.2292175	1.1008437	0.4907497	1.3435900
S5747	S5748					
0.4228581	1.3711343					

In this example, the Beta coefficient shows that S5735 tends to spread more his scores than the panel ($\text{Beta}=2.04 > 1$) whereas S5745 spreads less his scores than the panel ($\text{Beta}=0.49 < 1$). This can be illustrated by the graphic (Figure 1) of the regression plotted with the following command:

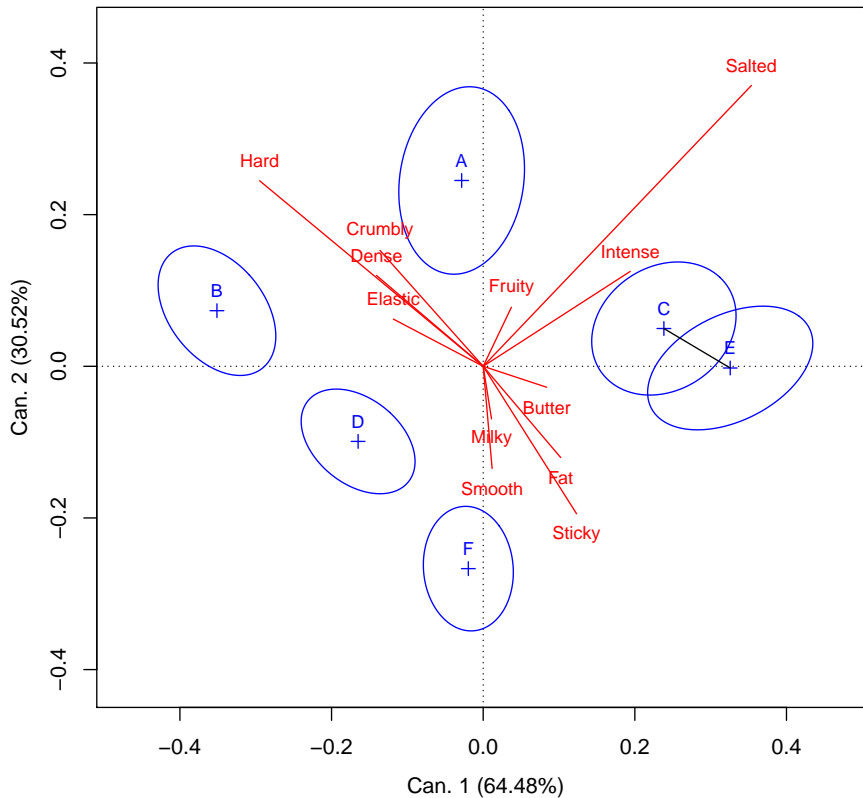
```
resOverall=MultiMAM(cheeses, modelType="overall", plotReg=TRUE)
```

The decomposition of the scores is obtained with `resMAM$decomposition` and the first line of the result is:

```
obs  ass  trt  X.mean  mu  centeredMean  assEffect  trtEffect
R1  S5733  A    1.9    3.575347  -1.675347  -0.9753472  -0.1586806
      int      err      Beta  scaling  disag      beta      va
-0.5746528  0.0333333  1.507171  -0.0804781  -0.4941747  1.507171  Butter
```

The first line of the decomposition checks $X.mean = \mu + assEffect + trtEffect + int + err$

Figure 2: Example of classical CVA
CVA – classic (94.99%)



NDIMSIG=2, Hotelling Lawley stat=7.808, F=8.149, pval=0, confidence ellipses at 0.9

and that $\text{int} = \text{disag} + \text{scaling}$ where $\text{scaling} = (\text{Beta}-1) * \text{trtEffect}$

3.3. MAM-CVA

We plotted the graphics of "tw" and "twe" CVA biplot (with the two-way model) in order to see the differences between classical CVA and MAM-CVA (Figure 3)

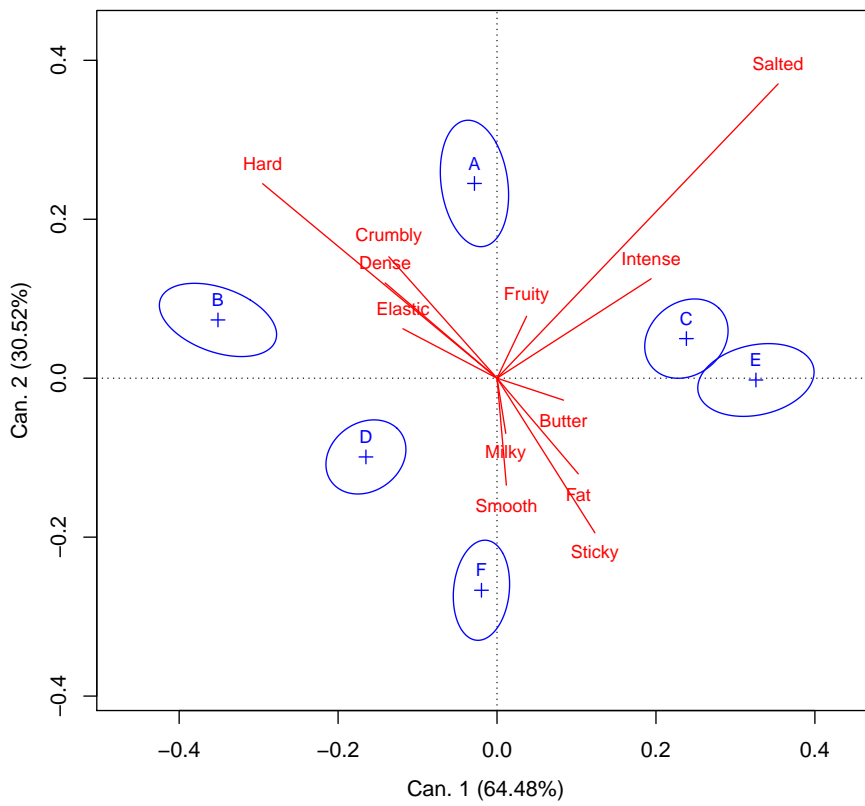
```
resCVATW=MAMCVA(cheeses, modelType="tw")
PlotCVA(resCVATW,mainTitle="CVA - classic")
resCVATWE=MAMCVA(cheeses,modelType="twe")
PlotCVA(resCVATWE,mainTitle="CVA - without panelist effect in ellipses")
```

As expected, the ellipses are consequently smaller in the right graph than in the left graph. The reduction of size is such as the Hotelling test found differences between products C and E (no segment links C and E on the right graph).

The MAM-CVA biplots (overall) was plotted on Figure 4.

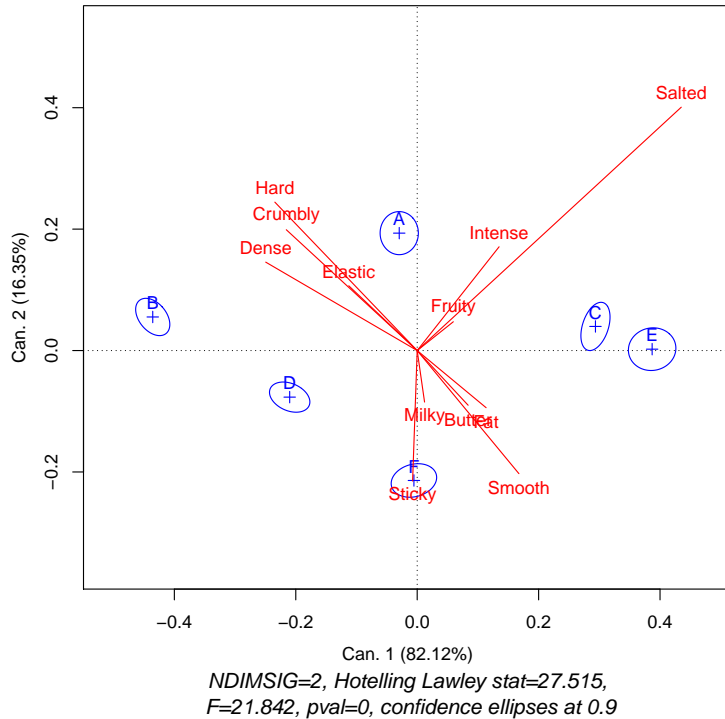
```
resCVAOverall=MAMCVA(cheeses, modelType="overall")
PlotCVA(resCVAOverall,mainTitle="MAM-CVA - overall")
```

Figure 3: Example of CVA without panelist effect in the ellipses (modelType="twe")
CVA – without panelist effect in ellipses (94.99%)



NDIMSIG=2, Hotelling Lawley stat=7.808, F=8.149, pval=0, confidence ellipses at 0.9

Figure 4: MAM-CVA (modelType="overall")
CVA (98.47%)



```
resCVAMAM=MAMCVA(cheeses, modelType="mam")
PlotCVA(resCVAMAM,returnX=TRUE,returnY=TRUE,mainTitle="MAM-CVA")
```

The parameters `returnX` and `returnY` allow the X and/or Y axes to be returned in order to have comparable maps.

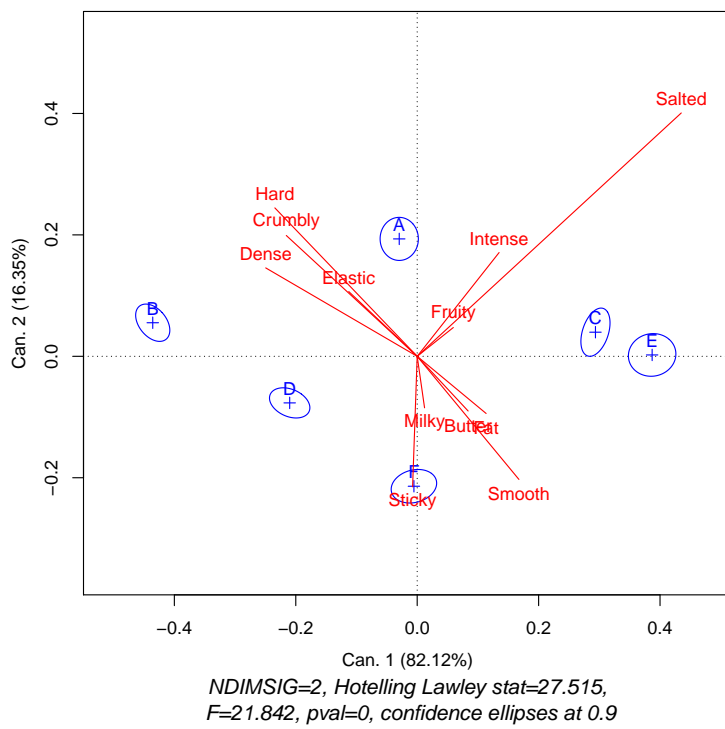
Slight differences between product and attribute positions exist between CVA and MAM CVA (Figure 4, Figure 5).

Regarding the analysis of the footnotes of the graphs, the F statistics were 8.14 for the two and two 2 CVA, 21 for the Overall MAM-CVA and 43 for the MAM-CVA. As expected, the F increases when the disagreement term decreases. The pvalue are lower than 0.001 and the number of significant dimensions is 2 for all methods.

Overall MAM-CVA biplot shows smaller ellipses than the classic and classic 2 CVA; and MAM CVA biplot shows smaller ellipses than overall MAM-CVA biplot.

We recommend to use the Overall MAM-CVA, even if the MAM-CVA is more discriminative. Indeed, the MAM-CVA corrects scaling effect for each descriptor separately whereas MAM-CVA corrects only scaling effect when it is observed on most of the descriptors. Removing assessor and scaling effects from the scores significantly decreases the size of ellipses. Furthermore, this approach is in agreement with the calculation of the F and with the Hotelling table.

Figure 5: MAM-CVA (modelType="mam")
CVA (98.47%)



4. Conclusion

Finally, MultiMAM allows the scores to be decomposed by taking scaling effect into account and plotting a map according to it. Two options are differentiated (i) the "mam" and (ii) the "overall" option. We advice to use the "overall" option in order to not "over-correct" the scaling effect. With this option, scaling is corrected by taking all descriptors into account. Furthermore, the overall scaling coefficient gives an indication of the psychological trend of an assessor to spread his scores. Ellipses representing the pure disagreement are smaller in MAM-CVA than in CVA, and more differences between products are noticed.

The **MultMAM** package takes a dataset as input and returns the decomposition of the scores and the map related to this decomposition. This work was presented with an example in sensory analysis. However, this model (and consequently this package) could also be valuable for any other field implying assessors scoring several items on different variables.

5. Acknowledgements

We are grateful to the Region of Burgundy which funded our researches with the JCE grant.

Bibliography

- Brockhoff PB, Schlich P, Skovgaard IM (2015). "Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model." *Food Quality and Preference*.
- Brockhoff PB, Skovgaard IM (1994). "Modelling individual differences between assessors in sensory evaluations." *Food Quality and Preference*, **5**(3), 215–224.
- Escoufier Y (1973). "Le traitement des variables vectorielles." *Biometrics*, **29**(4), 751–760.
- Mardia KV; Kent JBJ (1989). *Multivariate Analysis*. Probability and mathematical statistics.
- Naes T (1990). "Handling individual differences between assessors in sensory profiling." *Food Quality and Preference*, **2**(3), 187–199.
- Peltier C, Brockhoff PB, Visalli M, Schlich P (2014). "The MAM-CAP table: A new tool for monitoring panel performances." *Food Quality and Preference*, **32**, Part A(0), 24–27.
- Peltier C, Visalli M, Schlich P (2015a). "Canonical Variate Analysis of sensory profiling data." *Journal Of Sensory Studies*, (Accepted).
- Peltier C, Visalli M, Schlich P (2015b). "Multiplicative decomposition of the scaling effect in the Mixed Assessor Model into a descriptor-specific and an overall coefficients." *Food Quality and Preference*, (In review).
- Saporta G HG (1986). *Régions de confiance en analyse factorielle*, volume IV of *Data Analysis and Informatics*. Elsevier Science Publisher B.V.

A. Commenting the MultiMAM function

The MultiMAM function returns the decomposition of scores obtained when MultiMAM is applied to a dataset. It also returns the scaling coefficients, matrices of the different effects, number of negative scaling coefficients and significance of the scaling effect.

A.1. Parameters of the function

- **frame**: a data frame whose columns are (in the same order): "ass", "trt", "obs", "descriptor 1", ..., "descriptor d".
- **modelType="overall"**: a string indicating the type of the model, can be "overall" for the overall scaling, "mam" for the MAM scaling, "tw" for the classical two-way model
- **negativeCorrection=TRUE**: if TRUE, when a scaling coefficient is negative, no scaling effect is corrected. It corresponds to the adjusted model in [Brockhoff *et al.* \(2015\)](#)
- **correctOnlyIfSignificant=TRUE**: if TRUE, the scaling will be corrected only when significant at the panel level. It corresponds to the conditionnal model in the MAM paper ([Brockhoff *et al.* 2015](#)).
- **limitOfSignificance=0.2**: limit of significance used with the correct OnlyIfSignificant option.
- **plotReg=FALSE**: if TRUE, the graphics representing the linear regression related to the obtention of the overall beta are plotted. The abscisses are the centered panel scores and the ordinates are the centered panelist scores, each point correspond to one product and each color to one descriptor.

A.2. Output of the function

The output of the MultiMAM function is a list containing:

- **decomposition**
a dataframe with IJKD lines with columns as:
 - "obs": for the observation (replicate in sensory profiling),
 - "ass": for the assessor,
 - "trt": for the treatment (product in sensory profiling),
 - "va": for the variable,
 - X.mean: the score,
 - mu: the μ^d ,

- centeredMean: $y_{ijk}^d - \mu^d$,
- prodEffect: $y_{i..}^d - y_{...}^d$,
- assEffect: $y_{.j.}^d - y_{...}^d$,
- int $y_{ij.}^d - y_{i..}^d - y_{.j.}^d + y_{...}^d$,
- err: $y_{ijk}^d - y_{ij.}^d$,
- beta: scaling coefficient(MAM, overall or 1 for the classical model), corresponds to $\beta_i + 1 = \hat{\beta}_i$,
- scaling: $(\beta - 1)*prodEffect$,
- disag: int - scaling.

Thus, each observation is decomposed according to the MultiMAM model and the contribution of each effect is estimated.

- **beta**: scaling coefficient. It is a matrix when `modelType="mam"`, a vector when `modelType="overall"` and is null for other `modelType`.
- **avgMat**: vector of the μ^d .
- **trtMat**: matrix with D columns and I lines containing the averages of treatments (products in sensory profiling) $y_{.j.}^d$.
- **assMat**: matrix with D columns and J lines $y_{i..}^d - y_{...}^d$.
- **scalMat** matrix with D columns and I*J lines containing the scaling term of the decomposition for each pair `trt*assessor`.
- **disagMat** matrix with D columns and I*J lines containing the pure disagreement term of the decomposition for each pair `trt*assessor`.
- **errMat** matrix with D columns and I*J*K lines containing the error term of the decomposition.
- **intMat** matrix with D columns and I*J lines containing the interaction term of the decomposition.
- **nNeg**: matrix with I columns and 1 line if "overall" is chosen, D lines if "MAM" is chosen; containing the number of negative scaling coefficients.
- **pvalScal**: pvalue for the significance of scaling effect for the whole set of assessors.

B. Commenting the MAM-CVA function

In order to obtain a map, the first step is to obtain a CVA object. Then, the graph is plotted with this CVA object as input. This strategy is also present in the FactoMineR package with the PCA function. First, we describe the MAM-CVA of a dataset function, giving the results of the different calculations without plotting a map.

B.1. Parameters of the function CVA

- **frame**: the dataframe with colnames specified as `assName`, `trtName`, `obsName`, descriptor 1,..., descriptor D
- **option**: the chosen model as a string: "overall" for the Overall MultiMAM, "mam" for the MultiMAM, "tw" for the classical model, and "twe" for the classical model with ellipses taking not assessor effect into account.
- **test="Hotelling-Lawley"**: test for the Manova: a string in "Hotelling-Lawley", "Pillai", "Roy" or "Wilks"
- **nbAxes="auto"**: number of axes to be plotted in the CVA. When "auto" is chosen, `nbAxes` is the number of significant dimensions
- **hTable=TRUE**: if TRUE, the Hotelling table is returned as a matrix.
- **alpha=0.1**: limit for the test of dimensionality
- **trtName="ProductCode"**: name of the column containing the treatments (products in sensory profiling)
- **assName="SubjectCode"**: name of the column containing the assessors
- **obsName="Replicate"**: name of the column containing the observations (replicates in sensory profiling)
- **representation="biplot"**: a string. If "biplot", the coordinates of the products and attributes are calculated for a biplot map. If "twoMaps", the coordinates are calculated for two maps

B.2. Output of the function

This function returns a list containing

- **IndivCoord**: a matrix containing the coordinates of the products
- **VarCoord**: a matrix containing the coordinates of the descriptors
- **NbDimSig**: number of significant dimensions
- **HotellingTable**: hotelling table on the number of significant dimensions

- **ConditioningOfW**: conditioning of the W matrix (if <30 the matrix is well conditioned)
- **B**: covariance matrix of product effect
- **W**: covariance matrix of disagreement effect
- **EigenVectors**: eigen vectors of $W^{-1}B$
- **EigenValues**: eigen values of $W^{-1}B$
- **Stats**: list of statistics containing the MANOVA statistic as "stat", the related F ("F") and the related pvalue ("pval")
- **decomposition**:
a dataframe with IJKD lines with columns as:
 - "obs": for the observation (replicate in sensory profiling),
 - "ass": for the assessor,
 - "trt": for the treatment (product in sensory profiling),
 - "va": for the studied variables (descriptors in sensory profiling),
 - X.mean: the score related to this observation,
 - mu: the μ^d ,
 - centeredMean: $y_{ijk}^d - mu^d$,
 - trtEffect: $y_{i..}^d - y_{...}^d$,
 - assEffect: $y_{.j.}^d - y_{...}^d$,
 - int $y_{ij.}^d - y_{i..}^d - y_{.j.}^d + y_{...}^d$,
 - err: $y_{ijk}^d - y_{ij.}^d$,
 - beta: scaling coefficient(MAM, overall or 1 for the classical model), corresponds to $\beta_i + 1 = \tilde{\beta}_i$,
 - scaling: $(\beta - 1)*trtEffect$,
 - disag: int - scaling.
- **CenteredProductSubjectTable**: the matrix used to calculate the confidence ellipses
- **nbAxes**: number of axes to be plotted
- **wDemi**: the matrix used in the calculation of a CVA biplot
- **option**: string giving the chosen option ("mam", "overall", "tw" or "twe")
- **representation**: the chosen representation

Note that if option="tw", MAMCVA is the same than the CVA presented in [Peltier et al. \(2015a\)](#)

C. Commenting the PlotCVA function

PlotCVA plots the map related to a MAMCVA result.

C.1. Parameters of the function

- `res.CVA`: result of the function CVA (list)
- `assessors`: "none", "points" or "label". Allows the assessor scores to be projected on the map as points or labels
- `confInt=0.9`: limit of the confidence ellipses
- `ellipsesType="barycentric"`: string. If "barycentric" is chosen, the ellipses represent a zone where the mean point should be with a probability of `confInt`. If "individual" is chosen, the ellipses represent a zone where the panelistes should be with a probability of `confInt`.
- `productColors=NULL`: a vector allowing the color of the products to be chosen.
- `returnX=FALSE`: If TRUE, the X-axis is reversed.
- `returnY=FALSE`: If TRUE, the Y-axis is reversed.
- `linkBetweenProducts=TRUE`: If TRUE, the products statistically not different according to the Hotelling table are linked with a segment.
- `type="R"`: string. If "R", the graphic is plotted as an R graph. If "png", "jpg", "pdf" or "wmf", the graphic is saved in the working directory as a png, jpeg, pdf or wmf file.
- `fileName="CVA"`: string. If `type!="R"`, `fileName` is the name of the file created in the working directory.
- `mainTitle="CVA"`: string. Title of the graph

C.2. Output of the function

This function returns the graphic related to `res.CVA` and which is possibly saved in the working directory.

Affiliation:

Caroline Peltier
Centre des Sciences du Goût et de l'Alimentation
9E boulevard Jeanne d'Arc
21000 Dijon, France
E-mail: caroline.peltier@dijon.inra.fr

Journal of Statistical Software
published by the American Statistical Association
Volume VV, Issue II
MMMMMM YYYY

<http://www.jstatsoft.org/>
<http://www.amstat.org/>
Submitted: yyyy-mm-dd
Accepted: yyyy-mm-dd

Do we need to replicate in sensory profiling studies?

C. Peltier^{1,2,3}, N. Mammasse^{1,2,3}, M. Visalli^{1,2,3}, S. Cordelle^{1,2,3}, P. Schlich^{1,2,3}

¹CNRS, UMR6265 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

²INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France

³Université de Bourgogne, UMR Centre des Sciences du Goût et de l'Alimentation, F-21000

Dijon, France

E-mail : caroline.peltier@dijon.inra.fr

Abstract

Several evaluations are necessary to test individual repeatability in sensory profiling. But do additional replicates really provide more information than the first evaluation? Statistical analysis of 377 studies with at least two replicates showed that when using the first replicate alone:

- Significance of product discrimination was identical for 6 out of 7 attributes,
- 87% of the significant attributes had a similar vector of mean scores *ie* a correlation coefficient higher than 0.9,
- Significance of multivariate product discrimination assessed by MANOVA was identical
- 90% of the datasets had a similar product configuration (RV>0.9),
- the CVA product maps were similar (same product configurations, attribute positions...).

Furthermore, a third replicate seems reed really useless. Also, a high number of subjects or a small number of products in the study increased the similarities between the one-replicate datasets and the two replicate datasets. Consequently, once the panel is trained, do not seem to add more information to the interpretation.

Keywords: Sensory profiling, replication, correlation, ANOVA, MANOVA, RV, CVA, meta-analysis, SensoBase.

1. Introduction

Sensory profiling has been widely used in sensory analysis. The aim of the method is to describe the sensory characteristics of food products using a list of defined attributes. Panelists involved in a sensory profiling task are required to rate the perceived intensities of a number of attributes such as: sweetness, saltiness, crunchiness, etc. A classical sensory descriptive profile consists of two main phases: the training phase and the testing phase (also known as the measurement phase).

During these phases, the panelists often evaluate the set of products more than once. Replicates are then considered as independent repetitions of the experiment under identical experimental conditions. However, adding replicates involves some practical considerations, such as the potential differences within replicated samples, the choice of the number of samples to be tested by assessors without causing sensory fatigue, or the increase of the financial cost of the study.

39 In practice, several replicate structures can be found. Replicates can be randomized or balanced
40 within the full experiment (Williams latin square for example), and then no replicate or session
41 effects can be tested. It is also possible to serve the same products in separated tasting sessions,
42 possibly with several days in between. In this case, if product stability cannot be guaranteed,
43 adding replicates may not be realistic. Another type of replicate structure introduces product
44 units in a structured way: for example, samples can come from different batches. In the latter
45 two structures, care should be taken with the analysis of data. Replicate should be treated as a
46 random effect in the ANOVA model.

47 Most papers dealing with sensory profiling studies mention the use of two or three replicates.
48 Stone and Sidel (2004) suggest that, empirically, within-subject variability decreases somewhat
49 from the first to the second replicate, levelling at the third one, and that subsequent replicates
50 exhibit no further decrease in variability. The authors presume that two to three replicates are
51 adequate for a descriptive sensory profile. The need to replicate descriptive evaluations has also
52 been discussed in Moskowitz et al. (2008). The authors stated that the appropriateness of
53 replicates depends on the objective.

54 Replicates allow panel performance to be tested. Moskowitz (2008) and Munoz (2008) argued
55 that if a panel is well trained, there will be very little difference between replicates. Hence, for
56 sensory profiling studies, replicates may be desirable but they are never essential. In the same
57 discussion paper, Gacula (2008) argued that the maximum amount of information came with
58 the first judgment of each person. Hence, each additional judgment adds proportionally less
59 information.

60 Learning from actual data is essential in order to improve sensory methodologies and to
61 investigate different points of view, such as the usefulness of replicate in sensory profiling. In
62 this paper, we will examine the impact of replicates on statistical outputs using several sensory
63 profile studies from the SensoBase. The key question in our research is: would replicates change
64 the diagnoses of the studies regarding mapping, discrimination or mean scores corresponding to
65 no replicate?

66 For each study, univariate and multivariate statistical analyses were carried out in order to
67 achieve comparisons between the full dataset (including the first two replicates) and the
68 truncated dataset (subset of the data including the first replicate only).

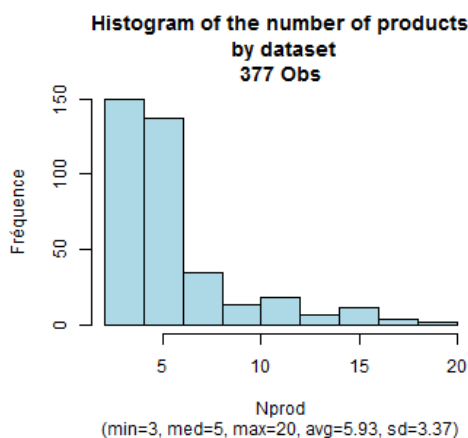
69 The same methodology was used to attest differences between replicates in datasets with a
70 small (then medium, and large) number of products, and in datasets with a small (then medium
71 and large) number of panelists. Furthermore, when a dataset contain three replicates, the
72 differences between the dataset constituted of the first two replicates and the whole dataset
73 were studied.

74 **2. Material and methods**

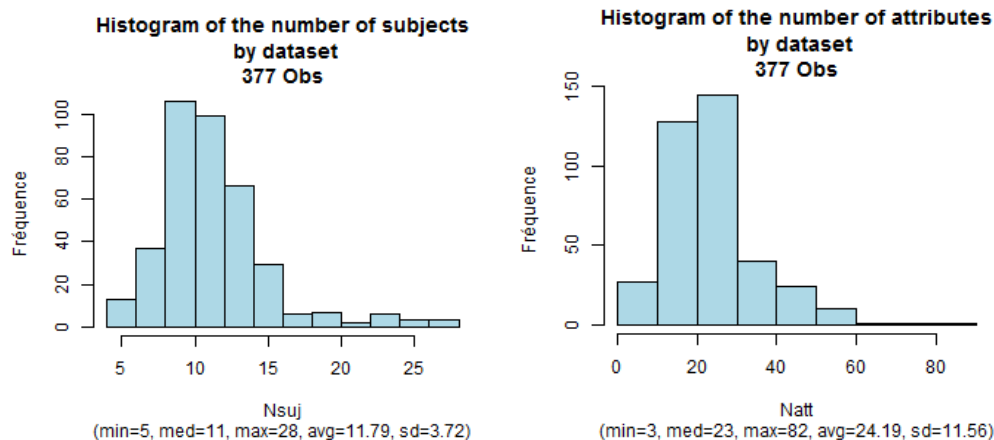
75 **2.1 Procedure**

76 SensoBase is a database of descriptive sensory studies, currently containing 1 169 datasets.
77 Each dataset contains the scores of a set of products of a given type profiled by a panel of
78 assessors. The database was created in order to describe sensory analysis practices, to compare
79 several statistical methods of analysis and to benchmark panel and panelist performances.

80 377 sensory profile datasets with at least two replicates were selected from SensoBase (each
 81 dataset selected had to contain between 2 and 21 products, more than 4 subjects and less than
 82 100 attributes). These datasets were chosen fully balanced to avoid potential statistical or
 83 computational issues. The distributions of the number of subjects, products and attributes are



84 presented in Figure 1.



85
 86 The full datasets (with 2 replicates) and the truncated datasets (with one replicate only) were
 87 compared according to unidimensional and multidimensional criteria. Univariate criteria
 88 (including discrimination and agreement between these datasets) were studied with two
 89 complementary approaches: by dataset (averaging on attributes) or by attribute (without taking
 90 any dataset effect into account). Multivariate criteria included multivariate discrimination,
 91 multivariate agreement, dimensionality of the product space and maps.

92 2.2. Univariate analysis

93 2.2.1 Panel discrimination

94 2.2.1.1. Approach by attribute

95 A two-way analysis of variance (ANOVA) model with Product and Assessor as factors was carried
 96 out on the intensity scores for each attribute. The additive model was chosen for analyses with
 97 one replicate, whereas the mixed two-way model with interaction and random subject effect
 98 was chosen for analyses with more than one replicate as recommended by (Næs & Langsrud,
 99 1998).

100 The significance of the product effect was calculated to attest the discrimination of products.

101 Note that performing an ANOVA with the mixed two-way model with subject as a random effect
102 is equivalent to average the replicates and then performing an additive two-way ANOVA.

103 When a significant product effect was found ($p < 0.1$) for the full dataset but not for the
104 truncated dataset, the sensory modality of the attribute was studied. This distribution of the
105 sensory modalities for these attributes was compared to the total distribution of sensory
106 modality with a Chi2 test in order to determine whether replicates are more useful for some
107 sensory modalities than for others.

108 2.2.1.2. Approach by dataset

109 For each dataset, the percentages of significant attributes were calculated for the full dataset
110 (with two replicates) and the truncated dataset, then the distributions were plotted and
111 compared.

112 The difference of those percentages (**percentage of significant attributes for the full dataset**
113 **minus the percentage of significant attributes for the truncated dataset**) was also calculated
114 for each dataset. It provides one indicator of comparison of univariate discrimination per
115 dataset. The average of this indicator was calculated and given as **differences in univariate**
116 **discrimination**.

117 2.2.2 Agreement between replicates

118 2.2.2.1. Approach by attribute

119 For each attribute, the Pearson correlation coefficient was calculated between the vector of
120 product mean scores for the full dataset and the vector of product mean scores for the
121 truncated dataset. Its distribution was plotted separately for significant attributes and non-
122 significant attributes.

123 2.2.2.2. Approach by dataset

124 The average of these Pearson correlation coefficients was calculated by dataset in order to
125 provide one indicator of the comparison of univariate agreement per dataset. The average of
126 this indicator was calculated and given as **univariate agreement**.

127 2.3. Multivariate analysis

128 2.3.1. Agreement between replicates: RV-coefficient

129 For each study, all the attributes were simultaneously analysed in order to examine whether or
130 not the same product space is derived from the full dataset and from the truncated dataset.

131 The RV coefficient (Escoufier, 1973) provides a simple way to measure the relationship between
132 two product spaces generated by two different sets of variables (sensory attributes), with values
133 between 0 and 1, where 1 represents perfect similarity. After averaging over assessors, the RV
134 coefficient was computed to measure the similarity between the two product configurations.
135 The average of RV coefficients was calculated and given as **multivariate agreement**.

136 2.3.2. Panel discrimination

137 Multivariate analysis of variance (MANOVA) measures product discrimination when all sensory
138 attributes are considered simultaneously. For each dataset, the two-way mixed MANOVA model

139 was applied on the intensity scores over all the attributes. This analysis is based on the
140 calculation of a matrix of discrimination $A = W^{-1}B$ where B and W are the covariance matrices
141 of product and interaction effects. We used the F-test based on the Hotelling –Lawley trace (the
142 sum of all the eigenvalues of the matrix A). A significant MANOVA F-test indicates that the
143 products differ significantly in the space generated by the attributes.

144 Then the ratio of the F obtained for the full dataset and the F obtained for the truncated dataset
145 was calculated (noted F_{full}/F_{trunc} . A ratio higher than 1 means that the full dataset was more
146 discriminative than the truncated one. As this ratio could be very high for some datasets, the
147 median of the F-ratio distribution was chosen as the comparison indicator of the multivariate
148 discrimination for several datasets (instead of the average, very sensitive to extreme values).

149 2.3.3. Representation: Canonical Variate Analysis

150 Canonical Variate Analysis (CVA) was used to map the product means (Peltier, Visalli, & Schlich,
151 2015). The axes which maximize the distances between product means while minimizing
152 dispersion for individual assessments of each product are defined by the eigenvectors of A.
153 Thus, the resulting map represents the **product discrimination** whereas the PCA of the product
154 mean table gives a map representing the **product mean variability** (without taking the subject
155 agreement around the means into account).

156 The use of MANOVA and CVA for sensory profile data has been described in Lawless and
157 Heymann (2010); Monrozier and Danzart (2001); Porcherot and Schlich (2000); Schlich (2004).
158 Peltier et al. (2015) recommend its use for sensory profiling data and provide an R package to
159 allow any sensory analyst to plot CVA maps.

160 Peltier, Visalli and Schlich (b) (2015) provided a methodology for comparing CVA and PCA maps
161 (first factorial plan constituted of the first two axes), calculating:

- 162 - the RV coefficient between PCA and CVA product maps to compare the product
163 configurations,
- 164 - the angles between one PCA and CVA attributes (for an identical product configuration)
165 to compare the sensory interpretations of the maps: the maximum and average of these
166 angles summarize the differences in sensory interpretations.
- 167 - the multivariate comparison of product pairs with the Hotelling T2 test for PCA and CVA :
168 the percentages of product pairs with similar results for both methods was calculated.

169 These criteria were used to compare the CVA maps for truncated and the full datasets.

170 When several datasets are compared, the averages of these criteria were calculated and noted
171 **product configuration, maximum angle, average angle** and **percentage of similar Hotelling**
172 **diagnosis**

173 2.3.4. The number of significant dimensions

174 A likelihood ratio test is used to assess the number of significant sensory dimensions reflecting
175 the main differences between products. The difference between the number of significant
176 dimensions with the full data and the number of significant dimensions with the truncated data
177 was studied. The percentage of datasets with the same number of significant dimensions was
178 also studied and named **percentage of datasets with same number of significant dimensions**.

179 2.4. Study of the datasets with three replicates

180 For datasets with three replicates, two comparisons were performed: (i) the comparison of the
181 truncated dataset (first replicates only) with the full dataset (three replicates), (ii) the
182 comparison of the truncated dataset (with the first two replicates) and the full dataset. Thus,
183 the relevance of a third replicates was studied. These comparisons were computed with both
184 descriptor and dataset approaches.

185 2.5. Influence of the number of products and number of subjects in the study on the different
186 indicators

187 The comparison indicators by dataset (univariate discrimination and agreement, multivariate
188 discrimination and agreement, and mapping differences) were declined according to the
189 number of products (low, medium, high) and the number of subjects (low, medium, high) of the
190 datasets. The distribution of the number of products allows product quantities to be defined as
191 low (less than 5 products), medium (between 5 and 9 products) and high (10 products and
192 more). The distribution of the number of subjects allow subject levels to be defined as low (less
193 than 10 subjects), medium (between 10 and 14 subjects) and high (15 subjects and more).

194 All statistical analyses and computations were performed using R software.

195 **3. Results and discussion**

196 3.1. Univariate results

197 3.1.1 Univariate discrimination

198 3.1.1.1. Approach by descriptor

199 For truncated datasets, results of the two-way ANOVA model on intensity scores for the first
200 replicate led to a significant product effect ($p = 0.1$) for 59% of the attributes. When the second
201 replicate is considered, the percentage increases to 66% (+7%). Since the median number of
202 attributes in a dataset was 23, considering the second replicates would therefore add on
203 average between one and two significant attributes.

204 86% of the attributes had the same F diagnosis (significant or not) whatever the number or
205 replicates. 10% were significant considering the full dataset, but non-significant with the
206 truncated dataset and 4% were significant considering the truncated dataset but non-significant
207 with the truncated one.

208 Table 1 shows the distribution of attributes according to their sensory modality for the 9121
209 attributes (general distribution), and for the 949 attributes becoming significant by adding one
210 replicate (specific distribution). A Chi2 test showed that the distributions were significantly
211 different ($p < 0.01$). Flavour and aftertaste attributes are slightly more represented in the specific
212 distribution than in the general distribution, suggesting that replicating could be slightly more
213 useful to discriminate the products for flavor or aftertaste attributes than for others. On the
214 contrary, texture attributes are less represented in the specific distribution suggesting that
215 replicating could be less useful for the discrimination of texture attributes.

216 Table 1: Distribution of attributes according to their sensory modalities (percentage) for all
 217 attributes (general) and for the attributes becoming significant with the addition of one
 218 replicate (specific).

	N	appearance	smell	taste	texture	flavor	aftertaste	trigeminal	others
general	9121	10.6	16.9	17.6	17.4	27.4	5.8	3.3	1.1
specific	949	7.4	16.1	18.0	13.4	32.1	8.0	4.0	0.9

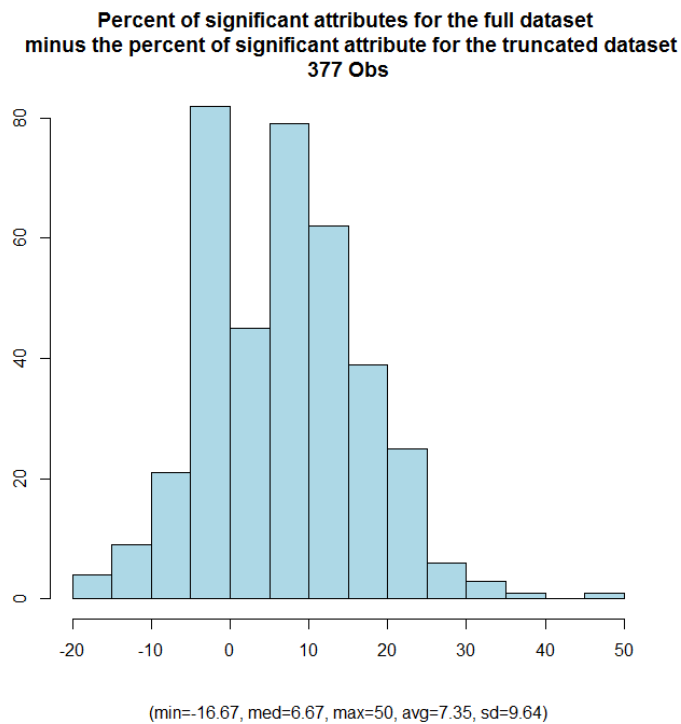
219

220 3.1.1.2. Approach by dataset

221 The distribution of the difference between the percent of significant attributes ($p < 0.1$) for the
 222 full datasets and the truncated ones is represented on Figure 2.

223 [INSERT FIGURE 2]

224 Figure 2: Histogram of the differences between the percent of significant attributes for the full
 225 dataset and the truncated one.



226

227

228 17.5% of the datasets have the same percentage of significant attributes in both cases. 13.3%
 229 of the datasets have more significant attributes by considering only the first rep, whereas 69.2%
 230 have more significant attributes by considering the full dataset.

231 3.2.2 Univariate agreement between the replicates

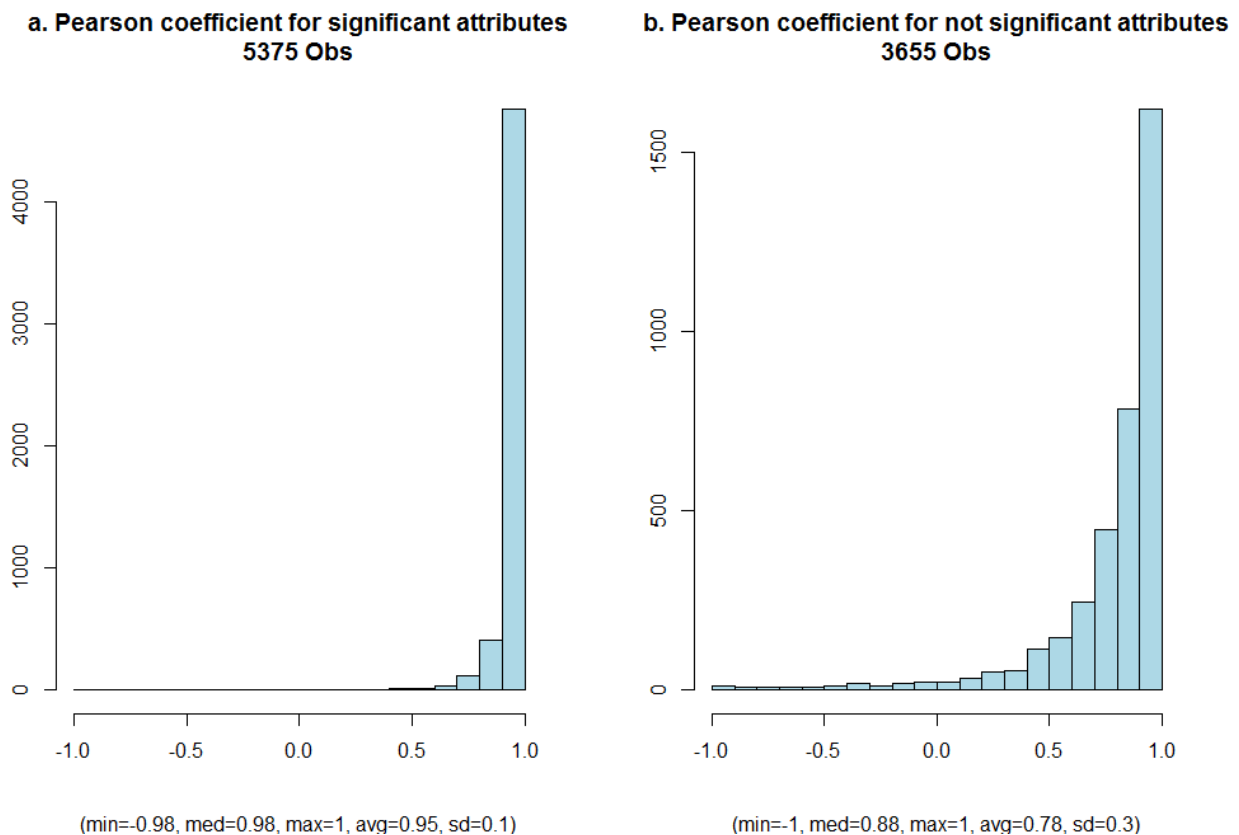
232 3.2.2.1. Approach by descriptor

233 87% of the significantly ($p = 0.1$) discriminative attributes had a correlation coefficient higher
234 than 0.9, denoting very good agreement between vectors of product mean scores, based on
235 one or two replicates. The mean correlation coefficient was 0.953.

236 [INSERT FIGURE 3]

237 Figure 3: Distribution of Pearson correlation coefficients between vectors of product mean
238 scores for discriminative (a) and non-discriminative (b) attributes of the full (2 reps) and
239 truncated (1 rep) datasets.

240

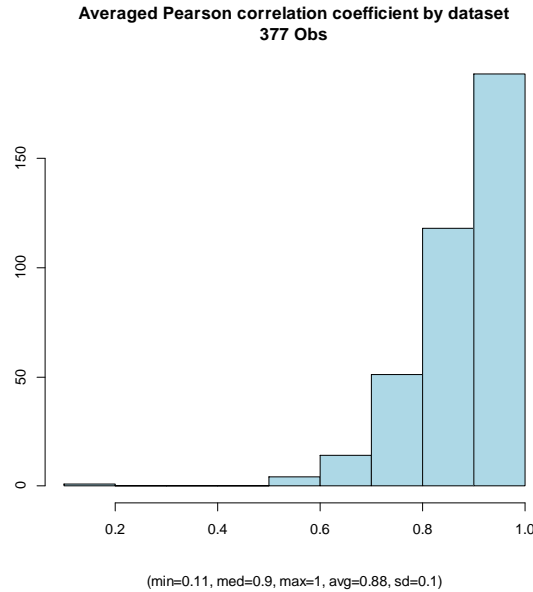


241 The distribution of the correlation coefficients (Figure 3) shows that 44% of the non-
242 discriminative attributes had a correlation coefficient higher than 0.9 and 9% lower than 0.45.
243 Since these attributes were non-discriminative, product mean scores would be closer with
244 changes in ranks leading to correlation coefficients lower than those obtained for discriminative
245 attributes. However the mean and median of these correlation coefficients were still correct
246 (0.778 and 0.878).
247

248 3.2.2.1. Approach by dataset

249 [INSERT FIGURE 4]

250 Figure 4: Distribution of averaged Pearson correlation coefficients by dataset between vectors
251 of product mean scores of the full (2 reps) and truncated (1 rep) datasets



252

253

254 The average of Pearson coefficient per dataset is higher than 0.9 for 50% of the datasets and
 255 higher than 0.8 for 81% of the datasets, showing an overall unidimensional agreement between
 256 the truncated and the full datasets, especially for the discriminated attributes.

257 3.2 Multivariate results

258 3.2.1 Multivariate discrimination

259 Table 2 shows that the gain in discrimination at a multivariate level was slightly enhanced after
 260 considering the second replicate. Indeed, the medians and means of the full dataset F were
 261 slightly higher than the truncated one. However, the improvement was not systematic, as
 262 shows the distribution of the F-ratio (when the F-ratio is lower than 1, the F was better by using
 263 one replicate only). Actually, 29.4% of the datasets had a better F with the truncated dataset
 264 than with the full dataset.

265 [INSERT TABLE 2]

266 Table 2: MANOVA F-statistic distribution for the full dataset and the truncated dataset

	Min	1st. Q.	Med	Mean	3rd. Q.	Max
Full dataset	1.458	3.79	6.402	16.37	11.93	1701
Truncated dataset	1.638	3.442	5.562	8.979	9.058	153.7
F_{full}/F_{trunc}	0.254	0.949	1.214	2.017	1.442	203.9

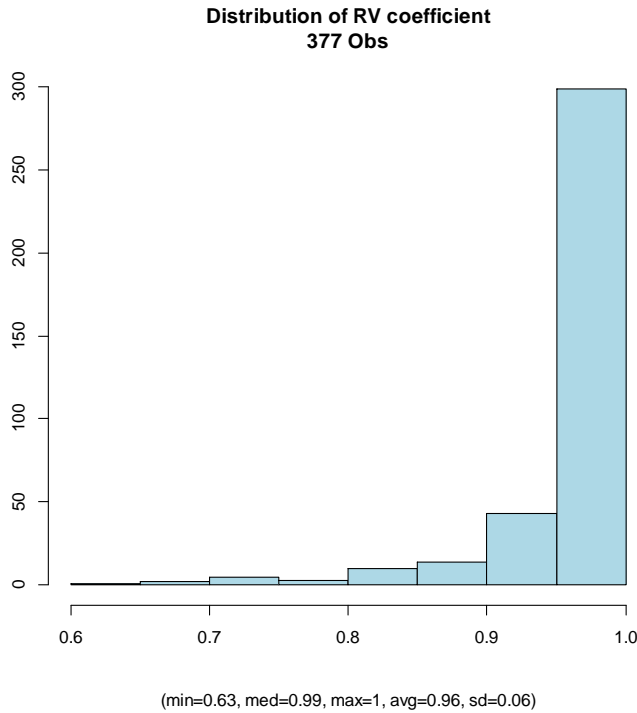
267 Note that the average of the distribution for the full dataset is artificially high because of
 268 extreme values.

269 3.2.2 Multivariate agreement: RV coefficient

270 The RV coefficient was higher than 0.9 for 90% of the datasets. Thus, the product configurations
 271 of the truncated and full datasets were almost the same.

272 [INSERT FIGURE 5]

273 Figure 5: Distribution of RV coefficients



274

275 3.2.3. CVA results

276 [INSERT TABLE 3]

277 Table 3: Comparison of the 1-2 map related to the full dataset and the 1-2 map related to the
 278 truncated dataset

	Number of datasets	Mean value
Product configuration	327	0.90
Maximum angle	22	36
Average angle	22	11
Percentage of similar Hotelling diagnosis	298	89%
Average of percentage of product pairs discriminated for the full dataset but not for the truncated dataset	298	5%

279 Table 3 shows that small differences exist between the map obtained with the full dataset and
 280 the truncated dataset: the magnitude of the differences is similar to thoses found between PCA
 281 and CVA in Peltier et al. (2014).

282 3.2.4. Number of dimensions

283 69% of the datasets had the same number of significant dimensions irrespective of the number
 284 of replicates. However, 22% had one or more supplementary significant dimensions by taking
 285 two replicates into account, whereas 9% had less significant dimensions by taking two
 286 replicates into account.

287 3.3. Study of the datasets with three replicates

288 3.3.1. Approach by descriptor

289 56% of the 1804 attributes were significant for both full (1-2-3 replicates) and truncated (1
 290 replicate) datasets, 15% for the full dataset only, and 3% for the first replicate only. The second
 291 and third replicates brought 12% of supplementary discriminant attributes.

292 66% of the 1804 attributes were significant for both full (1-2 -3 replicates) and truncated dataset
 293 (1-2 replicates), 5% for the full dataset and 3% for the first two replicates only. The third
 294 replicate did not bring much more discrimination.

295 3.3.2. Approach by dataset

296 According to table 6, dealing with only the first replicate gives results quite similar with the full
 297 dataset results, and dealing with the first two replicates gives results very similar with the full
 298 dataset results.

299 [INSERT TABLE 6]

300 Table 6: Comparison of the truncated dataset (first or first two replicates only) and the full
 301 dataset on 65 datasets

	1 VS 1-2-3 ¹	1-2 VS 1-2-3 ²
Differences in univariate discrimination (percentage of supplementary significant attributes for the full data)	12.23	2.53
Univariate agreement between replicates (Pearson coefficient average)	0.87	0.92
Multivariate discrimination (median of F-ratio F_{full}/F_{trunc})	1.19	1.12
Multivariate agreement (RV on product mean table)	0.95	0.98
Average of RV coefficient between the product configuration	0.86	0.96
Average of maximum angles	80	32
Average of average angles	18	7
Average of percentages of similar diagnosis (Hotelling T2)	86%	88%
Percentage of same number of sig. dimensions	57.4%	78.7%

302 ¹Comparison of the truncated dataset (first replicate only) and the full dataset (3 replicates)

303 ²Comparison of the truncated dataset (first two replicates only) and the full dataset (3
 304 replicates)

305 3.4. Influence of the number of products and number of subjects in the study on the different
 306 indicators

307 Table 7 shows that the differences between the full and the truncated datasets are higher for
 308 higher number of products. In other words, the more products, the more relevant a second
 309 replicate is.

310 [INSERT TABLE 7]

311 Table 7: Criterion of comparison depending on the number of products in the related dataset

	Low number of products <5	Medium number of products	High number of products >9
Number of datasets	150	179	48
Differences in univariate discrimination	6.24	7.83	9.00
Univariate agreement	0.84	0.90	0.86
Multivariate discrimination	1.17	1.23	1.21
Multivariate agreement	0.965	0.9695	0.94
Product configuration	0.91	0.89	0.89
Percentage of similar Hotelling diagnosis	88%	91%	84%
Maximum angle	19	41	60
Average angle	7	11	17
Percentage of datasets with the same number of significant dimensions	76%	72%	41%

312

313 Table 8 shows that the differences between the full and the truncated datasets are higher for
 314 smaller number of subjects. In other words, the more subjects, the less relevant a second
 315 replicate is.

316 [INSERT TABLE 8]

317 Table 8: Criterion of comparison depending on the number of subjects in the related dataset

	Low number of subjects <10	Middle number of subjects	High number of subjects (>14)
Number of datasets	86	165	56
Differences in univariate discrimination	8.14	7.12	7.7
Univariate agreement	0.865	0.876	0.886

Multivariate discrimination	1.17	1.20	1.29
Multivariate agreement	0.9607	0.9632	0.9605
Product configuration	0.86	0.91	0.92
Percentage of similar Hotelling diagnosis	87%	91%	91%
Maximum angle	54	40	24
Average angle	22	11	7
Percentage of datasets with the same number of significant dimensions	61%	74%	66%

318

319 **4. Conclusion**

320 In this work, the differences between statistical analysis on the dataset restricted to the first
321 replicate and the full dataset were quantified, showing minor differences.

322 Adding a second replicate tends to slightly improve the univariate and multivariate
323 discrimination (6 attributes over 7 have the same univariate discrimination with or without
324 replicates), and can increase the number of dimensions of the product space. Adding a third
325 replicate still improves (but very slightly) the univariate discrimination.

326 The stability of the statistical analysis seems to be linked to the number of products and
327 subjects: the more products and the less subjects, the more unstable the statistical analysis is.
328 Consequently, a solution to avoid replicates and keep the same accuracy could be to include
329 more subjects, or less products in the analysis.

330 Replicates are required to test the repeatability of the panel and panelists. Once the panelists
331 are trained, the need of replicates is not prominent. As replicates are time and budget
332 consuming, the panel leader should make a compromise between more accurate sensory data
333 and loss of time and money. Further work will investigate how many supplementary subjects
334 would be required to obtain the same results than those obtained by adding a replicate.

335

336 **References**

337 Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29(4):751–760.

338 Gacula, M. (2008). Replication in sensory and consumer testing. In *Viewpoints and*
339 *Controversies in Sensory Science and Consumer Product Testing*, pages 306–311. Food &
340 Nutrition Press, Inc.

341 Lawless, H. T. and Heymann, H. (2010). Sensory evaluation of food. pages 433 – 449. *Food*
342 *Science Text Series*.

343 Monrozier, R. and Danzart, M. (2001). A quality measurement for sensory profile analysis -the
344 contribution of extended cross-validation and resampling techniques. *Food Quality and*
345 *Preference*, 12(5-7):393–406.

346 Moskowitz, H. (2008). Replication in sensory and consumer testing. In *Viewpoints and*
347 *Controversies in Sensory Science and Consumer Product Testing*, pages 299–302. Food
348 & Nutrition Press, Inc.

349 Moskowitz, H., Munoz, A., and Gacula, M. (2008). Replication in sensory and consumer testing.
350 In *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*, pages 299–
351 311. Food & Nutrition Press, Inc.

352 Munoz, A. (2008). Replication in sensory and consumer testing. In *Viewpoints and Controversies*
353 *in Sensory Science and Consumer Product Testing*, pages 302 – 306. Food & Nutrition Press, Inc.

354 Naes, T., Langsrud, O. (1998). Fixed or random assessors in sensory profiling ? *Food Quality and*
355 *Preference*, 9(3), 145-152

356 Peltier, C., Visalli, M., Schlich, P. (2015), Comparison of Canonical Variate Analysis and Principal
357 Component Analysis on 422 descriptive sensory studies. *Food Quality and Preference*, 40B, 326-
358 333

359 Peltier, C., Visalli, M., Schlich, P. (2015) (b). Canonical Variate Analysis of sensory profiling data.
360 *Journal of Sensory Studies* (accepted)

361 Porcherot, C. and Schlich, P. (2000). Flash table and canonical mapping of potato varieties. *Food*
362 *Quality and Preference*, 11(1-2):163–165.

363 Schlich, P. (2004). L'analyse en variables canoniques de données de profils sensoriels. Rennes,
364 France. 8eme journees Europeennes Agro-industrie et Methodes Statistiques.

365 Stone, H. and Sidel, J. L. (2004). Test strategy and the design of experiments. In *technology, F. s.*
366 *a., editor, Sensory Evaluation Practices*, pages 116–117. Elsevier Academic Press, USA, third
367 edition.

368

369

370

ANNEXE 2 :

RAPPORTS DE META-ANALYSES

A. RAPPORT R1 : ACP CONTRE CVA, META-ANALYSE DES BIPLOTS

I. PROBLEMATIQUE

Jusqu'à présent, les différences entre ACP et CVA ont été mesurées sur des cartes séparées en deux graphiques, un pour les descripteurs, l'autre pour les produits. Il existe un autre type de carte permettant de représenter produits et descripteurs sur le même graphique : le **biplot** (Gower, 2011). Cette méthode n'est **pas** équivalente à la superposition des graphiques séparés, mais permet de projeter les produits sur les descripteurs. La question des différences entre les deux biplots se pose donc naturellement.

II. PROTOCOLE

1. ANALYSES STATISTIQUES MISES EN PLACE POUR REPONDRE A LA PROBLEMATIQUE

Les critères mis en place (Peltier, Visalli, & Schlich, 2015a) pour la comparaison ACP CVA ont été calculés pour cette méta-analyse :

- le coefficient RV entre les configurations produit (**RVCoef**),
- l'angle moyen entre les descripteurs (après superposition des configurations des produits) (**avgAngle**),
- l'angle maximal entre les descripteurs (après superposition des configurations des produits) (**maxAngle**),
- le ratio entre les F de la MANOVA de la CVA et l'ACP (**F-ratio**),
- la comparaison des intersections des ellipses de confiance : proportion de paires de produit au diagnostic identique (« les ellipses sont disjointes » ou « les ellipses se chevauchent ») pour les deux méthodes (**pId**),
- le pourcentage de paires de produits discriminées dans la CVA et pas dans l'ACP (**PctCVA sig**).

2. CARACTERISTIQUES DES JEUX DE DONNEES A SELECTIONNER DANS L'ANALYSE

Des jeux de données ont été sélectionnés de manière à éviter les jeux de données atypiques ou pour lesquels la CVA ne fonctionnerait pas. Les caractéristiques des jeux de données sélectionnés ont été les suivantes :

- jeux de données équilibrés (même nombre de répétitions par sujet pour chaque produit),
- plus de 2 et moins de 20 produits,
- moins de 100 descripteurs,
- inversion de matrice possible dans la CVA,
- un nombre de descripteurs significatifs supérieur à 2

III. PRESENTATION DE LA BASE DE DONNEES OBTENUE

La base de données obtenue est composée de 445 jeux de données et a les caractéristiques présentées dans le Tableau A. 1 et Figure A. 1.

Nombre de sujets	Nombre de produits	Nombre de descripteurs	Nombre d'observations
2 087	2 571	6 677	1 489 511

TABLEAU A. 1 : CARACTERISTIQUES DE LA BASE DE DONNEES OBTENUE

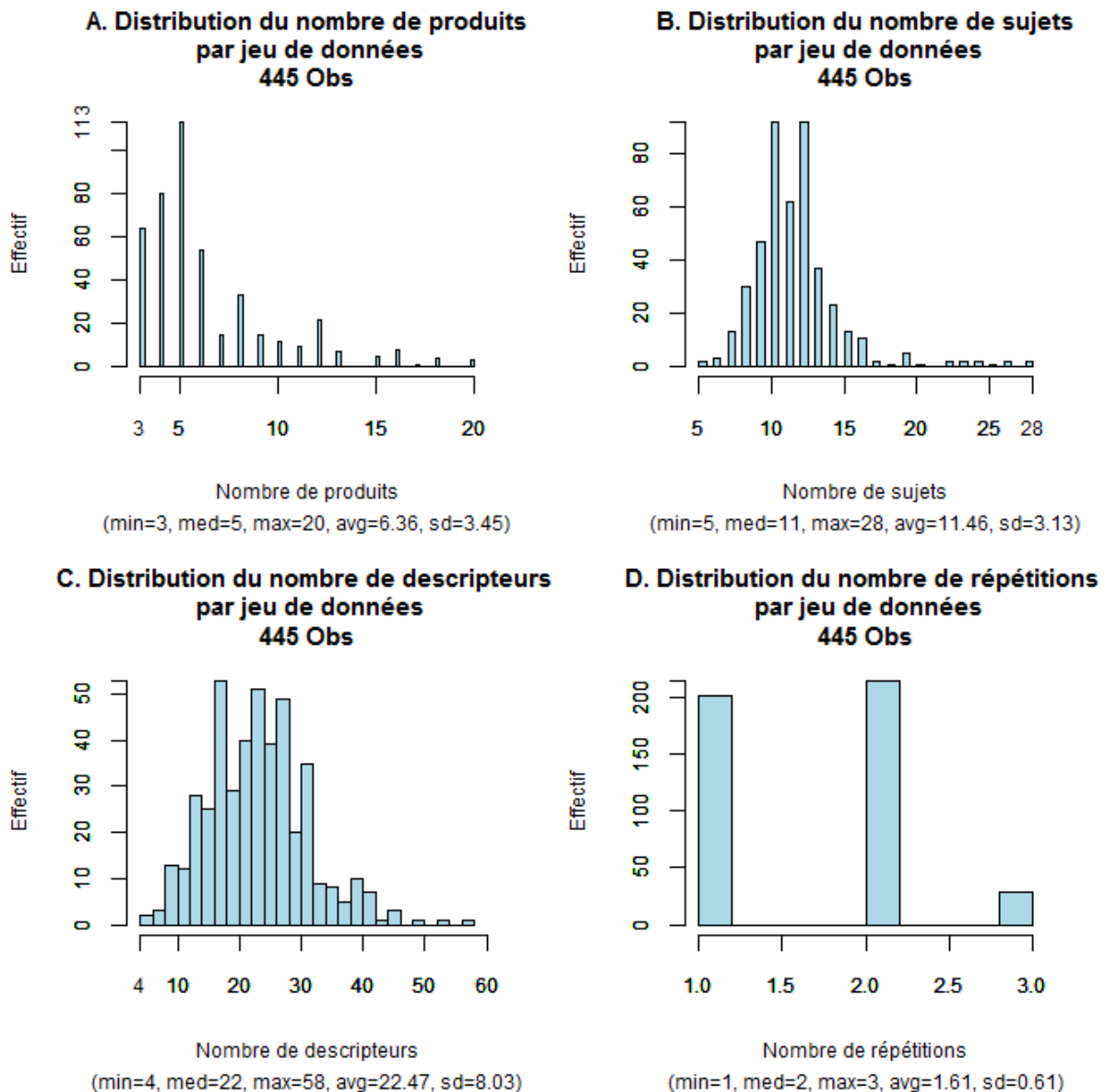


FIGURE A. 1 : DISTRIBUTIONS DES NOMBRES DE PRODUITS (A) SUJETS (B), DESCRIPTEURS (C) ET REPETITIONS (D) DES JEUX DE DONNEES UTILISES DANS LA META-ANALYSE DE COMPARAISON DES BIPLOTS DE L'ACP ET DE LA CVA

IV. RESULTATS DES ANALYSES STATISTIQUES

Les distributions des critères de comparaison sont présentées Figure A.2 et Tableau A. 2.

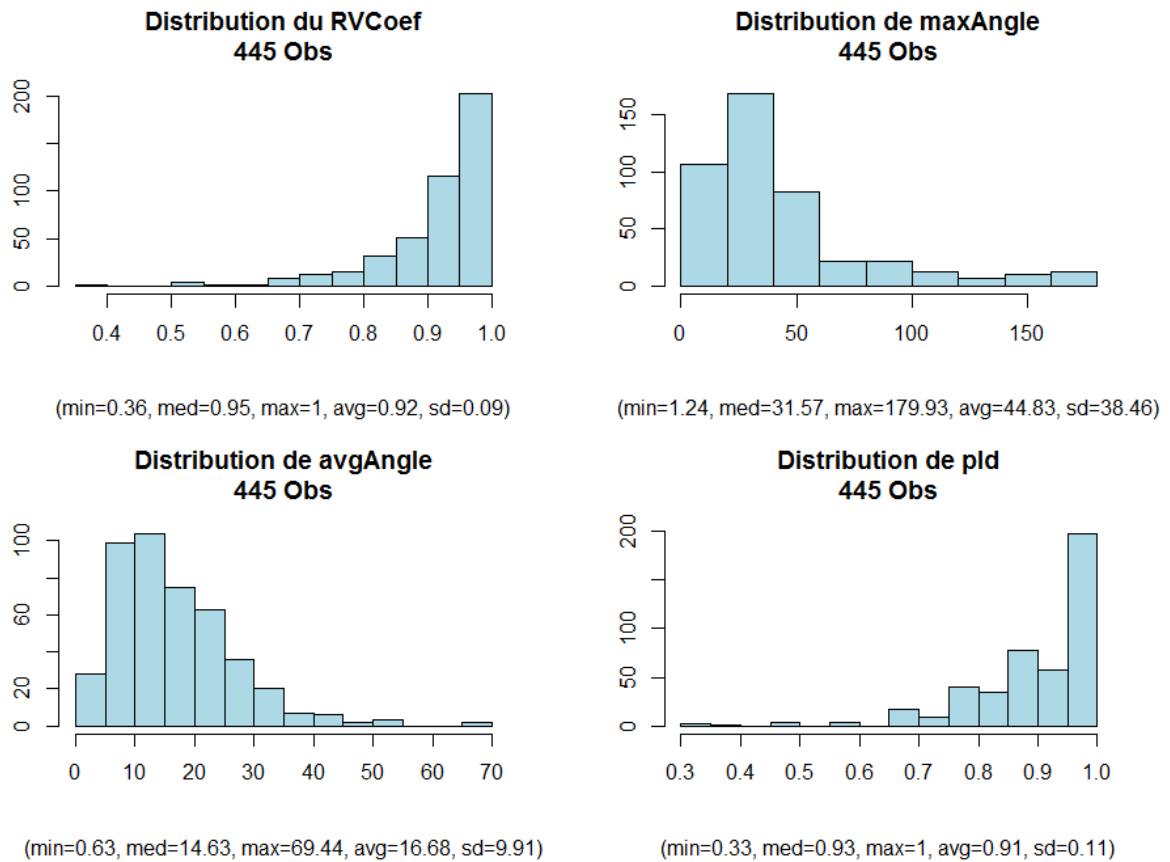


FIGURE A.2 : DISTRIBUTIONS DES CRITERES DE COMPARAISON DES BIPLOTS DE L'ACP ET DE LA CVA

	Min	1Q	Med	Mean	3rd Q	Max
Fratio	1.008	1.351	1.766	2.399	2.458	28.43
PctCVAsig	0	0	0	3.75%	6.7%	33.3%

TABLEAU A. 2 : DISTRIBUTIONS DU F-RATIO ET DU POURCENTAGE DE DESCRIPTEURS SIGNIFICATIFS SEULEMENT AVEC LA CVA

V. INTERPRETATION ET DISCUSSION

Les biplots d'ACP et de CVA semblent donner des cartes relativement semblables (configuration produit, discrimination...). Cependant, de légères différences de directions des descripteurs ont été observées (les différences de directions des descripteurs ont été de 16 degrés en moyenne, et l'angle maximal entre les directions des descripteurs a été de 44 degrés en moyenne).

Les différences observées entre les biplots ont été légèrement plus importantes que les différences entre les cartes d'ACP et de CVA « non-biplots » (c'est à dire avec un graphique représentant les produits et l'autre représentant les corrélations) (Peltier et al., 2015a).

B. RAPPORT R2 : COMPARAISON DE LA MAM-CVA ET DE L'ACP SUR 379 JEUX DE DONNEES DE LA SENSOBASE

I. PROBLEMATIQUE

La CVA est une carte prenant en compte la variabilité des sujets autour des moyennes des produits. Les ellipses de confiance de la CVA représentent la variabilité totale des notes autour des moyennes des produits. La CVA sans la projection de l'effet sujet dans les ellipses permet d'obtenir des ellipses plus petites, représentant uniquement l'interaction entre les sujets : l'effet sujet –c'est-à-dire le fait qu'un sujet note plus ou moins haut - a été retiré.

La MAM-CVA est une carte prenant en compte l'effet de scaling (Peltier, Visalli, & Schlich, 2015b) à l'aide d'un modèle multidimensionnel : le MultiMAM, décliné en deux versions : « global » (l'effet de scaling est commun à tous les descripteurs) ou « multivarié » (l'effet de scaling est spécifique à chaque descripteur). La problématique de cette étude est de comparer les cartes obtenues par ACP et par CVA sans effet sujet et MAM-CVA.

II. PROTOCOLE

1. ANALYSES STATISTIQUES MISES EN PLACE POUR REPONDRE A LA PROBLEMATIQUE

La CVA sans la projection de l'effet sujet et la MAM-CVA (global et multivariée) ont pour particularité de prendre en compte l'effet sujet (et l'effet de scaling pour la MAM-CVA) dans la construction des ellipses de confiance. L'ACP est toujours calculée sur le tableau des moyennes des produits (descripteurs en colonnes, produits en lignes).

Les critères de comparaison des cartographies (Peltier et al., 2015a) sont:

- le coefficient RV entre les configurations produit (**RV coef**),
- l'angle moyen entre les descripteurs (après superposition des configurations produits par rotation procrustéenne) (**avgAngle**),
- l'angle maximal entre les descripteurs (après superposition des configurations produits) (**maxAngle**),
- le ratio du F de la MANOVA pour la CVA et l'ACP (**F-ratio**),
- le pourcentage des paires de produits discriminées en MAM-CVA et pas en ACP, (**PctDiscrCVA**)
- la proportion de paires de produits avec un diagnostic identique pour la MAM-CVA et l'ACP (**pId**)

2. CARACTERISTIQUES DES JEUX DE DONNEES SELECTIONNES DANS L'ANALYSE

Des jeux de données ont été sélectionnés suivant les critères suivants :

- équilibrés,
- plus de 2 et moins de 20 produits,
- moins de 100 descripteurs,
- l'inversion de matrice possible dans la MAM-CVA et la CVA

- avec un nombre adéquat de descripteurs significatifs (supérieur à 2 et inférieur au nombre d'observations).

III. PRESENTATION DE LA BASE DE DONNEES OBTENUE (379 JEUX DE DONNEES)

Une base de données à 379 jeux de données ayant les caractéristiques suivantes est obtenue (Figure A 3 et Tableau A. 3).

Nombre de sujets	Nombre de produits	Nombre de descripteurs	Nombre d'observations
2 743	2 161	8 165	1 287 499

TABLEAU A. 3 : CARACTERISTIQUES DE LA BASE DE DONNEES UTILISEE POUR LA META-ANALYSE

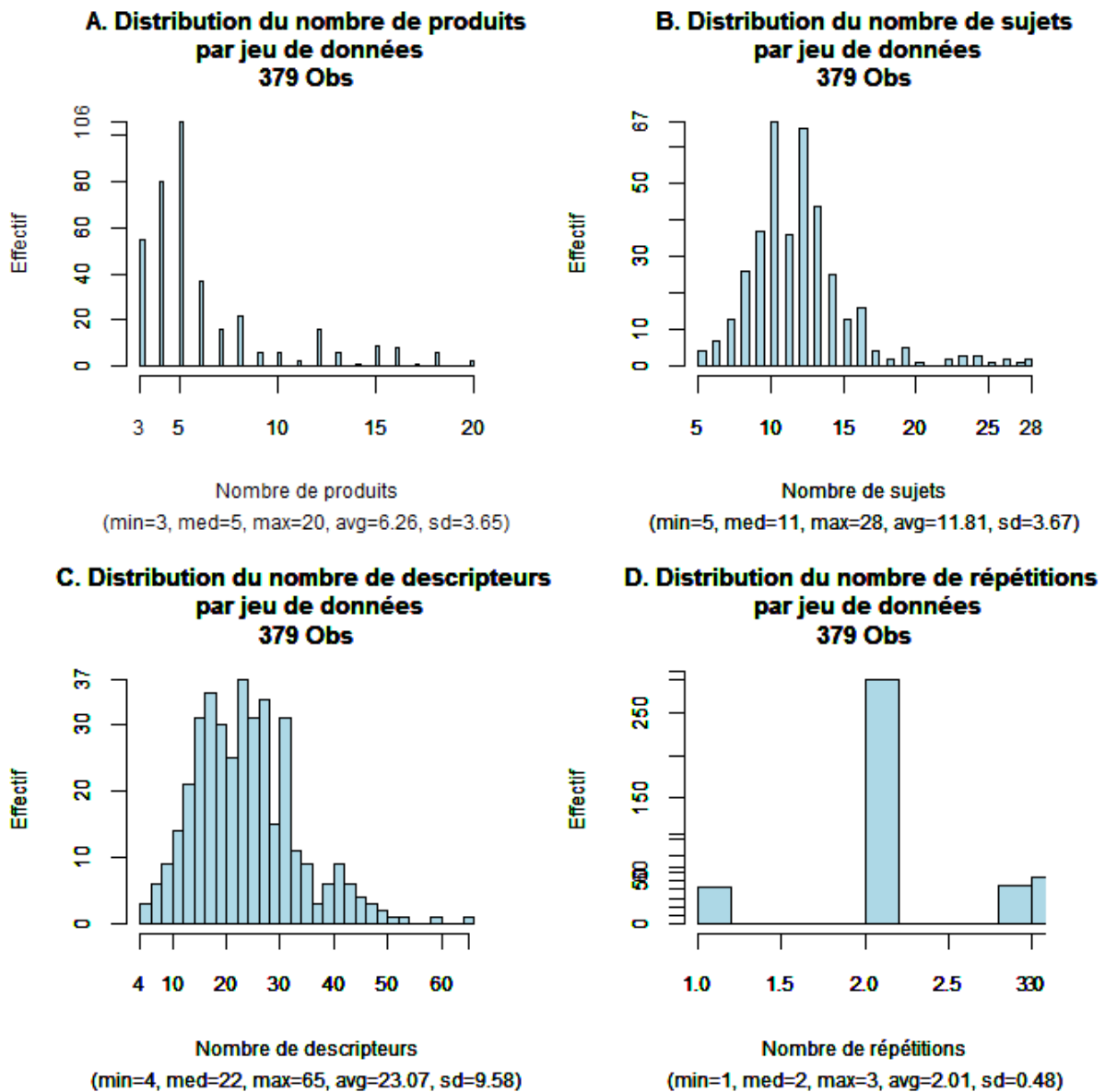


FIGURE A 3 DISTRIBUTIONS DES NOMBRES DE PRODUITS (A) SUJETS (B), DESCRIPTEURS (C) ET REPETITIONS (D) DES JEUX DE DONNEES DE LA META-ANALYSE ACP CONTRE CVA-MAM

IV. RESULTATS DES ANALYSES STATISTIQUES

Le Tableau A. 4 résume les différents critères de comparaison :

Critère de comparaison	de ACP/CVA (Peltier et al., 2015a)	ACP/MAM-CVA (global)	ACP/MAM-CVA(multivariée)	ACP/CVA sans projeter l'effet sujet
Nombre de jeux de données	422	379	377	379
RVcoef	0.91	0.95	0.92	0.92
MaxAngle	33	23	25	26
AvgAngle	6	6	7	7
pId	5%	0.3%	0.3%	0.6%
PctDiscrCVA	4%	18%	19%	17%
IdPct	91%	81%	80%	82%
F-ratio (med)	1.8	1.3	1.13	1.8

TABLEAU A. 4 : MOYENNES DES CRITERES DE COMPARAISON DE CARTES (ACP VS CVA, MAM CVA OU ADAPTED CVA)

V. INTERPRETATION ET COMMENTAIRES

Mis à part la taille des ellipses de confiances (et leurs intersections), les cartes de MAM-CVA semblent proches de celles de l'ACP (et, étonnamment, même légèrement plus que la CVA ne l'est elle-même). Finalement, la méthode choisie ne semble pas faire beaucoup varier les configurations des produits et les interprétations sensorielles. En revanche, lorsqu'on considère des ellipses qui prennent en compte l'effet de scaling et l'effet sujet, 15% de séparations d'ellipses (ou de différences entre produits) supplémentaires ont été observés. La séparation semble être essentiellement due au retrait de l'effet sujet. L'effet de scaling a permis d'ajouter 1% (en moyenne) de paires de produits discriminées dans le cas global et 2% pour le cas multivarié.

C. RAPPORT R3 : COMPARAISON DE L'ACP ET DE L'ACP VERTICALE PAR META-ANALYSE DE LA SENSOBASE

I. PROBLEMATIQUE

En profil sensoriel, l'ACP est basée sur la matrice des moyennes des produits pour l'ensemble des descripteurs (parfois restreinte aux descripteurs significatifs uniquement). Le plus souvent, cette matrice comporte plus de variables (descripteurs) que d'observations (produits), ce qui représente un problème. Pour le pallier, Dettmar, Plattner, and Vetter (2014) propose une méthode d'ACP : l'ACP Verticale. Cette méthode est basée, non plus sur la matrice des moyennes des produits, mais sur une matrice contenant autant de lignes qu'il y a eu d'observations durant l'expérimentation (produit du nombre de répétitions, du nombre de produits et du nombre de sujets). L'objectif de cette étude est de comparer les résultats donnés par l'ACP et l'ACP Verticale et de vérifier l'hypothèse avancée par (Dettmar et al., 2014) qui suggère que l'ACP Verticale doit apporter plus de dimensionnalité à l'espace produit.

II. PROTOCOLE

1. ANALYSES STATISTIQUES

L'ACP Verticale est calculée comme suit :

Les données du tableau donnant les notes pour chaque produit, sujet et répétition sont centrées par sujet. Elles sont ensuite divisées par l'écart type du sujet pour donner le même poids à chaque sujet. Cet écart type a été obtenu en calculant la variance de chaque descripteur, puis en calculant la moyenne de ces variances sur l'ensemble des descripteurs, pour finalement en prendre la racine.

La table résultante est ensuite diagonalisée.

Ainsi, chaque ACP verticale a été comparée à l'ACP à l'aide des critères définis dans Peltier et al. (2015a):

- coefficient RV entre les configurations produit (**RV coef**),
- angle moyen entre les descripteurs (après superposition des configurations produits) (**avgAngle**),
- angle maximal entre les descripteurs (après superposition des configurations produits) (**maxAngle**),
- ratio du F de la MANOVA pour la ACP et l'ACP Verticale (**F-ratio**),
- pourcentage des paires de produits discriminées en MAM-CVA et pas en ACP, (**PctDiscrCVA**)
- proportion des paires de produits avec diagnostic identique pour la MAM-CVA et l'ACP (**pId**)

De plus, deux critères de dimensionnalité ont été calculés pour chaque jeu de données :

- Le coefficient Beta = $1 + 2 \frac{\sum_{i < j}^N \lambda_i \lambda_j}{\sum_i^N \lambda_i^2}$, où λ_i sont les valeurs propres de l'ACP. Ce coefficient a été proposé par Schlich (1996) pour les données de profil sensoriel
- N : nombre d'axes requis pour représenter 80% de la variance totale.

Comme le nombre de dimensions choisies dépend du nombre de produits, nous avons tout d'abord effectué cette étude de dimensionnalité sur l'ensemble des jeux de données puis nous l'avons restreint aux jeux de données dans lesquels le nombre de produits était plus grand que le nombre de descripteurs significatifs.

De la même manière, l'ACP des corrélations et l'ACP Verticale seront comparées. De plus, on reportera les résultats obtenus dans Peltier et al. (2015a) pour comparer l'ACP à la CVA et l'ACP des corrélations à l'ACP pour donner des ordres d'idées des différences obtenues.

2. SELECTION DES JEUX DE DONNEES

Nous avons sélectionné les jeux de données équilibrés ayant :

- au moins 3 produits,
- plus de 4 sujets,
- moins de 100 descripteurs,

III. PRESENTATION DE LA BASE DE DONNEES (411 JEUX DE DONNEES)

Une base de données à 411 jeux de données est obtenue avec les caractéristiques décrites dans le Tableau A. 5 et la Figure A. 4.

Nombre de sujets	Nombre de produits	Nombre de descripteurs	Nombre d'observations
2978	2293	9394	1439083

TABLEAU A. 5 : TABLEAU DESCRIPTIF DE LA META-ANALYSE DE COMPARAISON DE L'ACP ET DE L'ACP VERTICALE

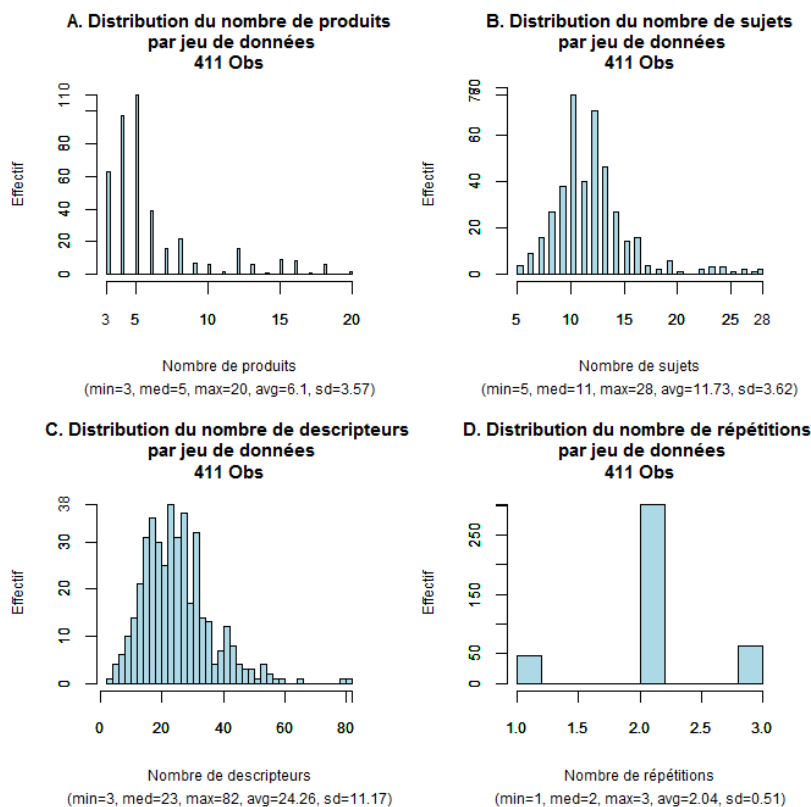


FIGURE A. 4 : DISTRIBUTIONS DES NOMBRES DE PRODUITS (A) SUJETS (B), DESCRIPTEURS (C) ET REPETITIONS (D) DES JEUX DE DONNEES LORS DE LA META-ANALYSE PORTANT SUR L'ACP ET L'ACP VERTICALE

IV. RESULTATS

Les résultats obtenus ont été résumés dans le Tableau A. 6, le Tableau A. 7, le Tableau A. 8 et le Tableau A. 9. Les distributions des indicateurs sont présentées Figure A. 5.

	ACP vs CVA	ACP vs ACP Vert
	1-2	1-2
RV coef	0.91	0.99
maxAngle	32.96	51
avgAngle	5.9	16
pId	0.91	0.94
Pourcentage de paires discriminées en ACP mais pas en ACP Verticale		5%
Pourcentage de paires discriminées en ACP Verticale mais pas en ACP		1%
F-ratio ACP/ACP Verticale		1.17

TABLEAU A. 6 : RESULTATS DE LA COMPARAISON ENTRE ACP ET ACP VERTICALE

	ACP vs ACP cor	ACP cor vs ACP Vert
	1-2	1-2
RV coef	0.92	0.92
maxAngle	28.79	57
avgAngle	7.10	17
pId	0.93	0.92
Pourcentage de paires discriminées en ACP mais pas en ACP Verticale		5.4%
Pourcentage de paires discriminées en ACP Verticale mais pas en ACP		2.7%
F-ratio ACP/ACP Verticale		1.06

TABLEAU A. 7 : RESULTATS DE LA COMPARAISON ENTRE ACP DES CORRELATIONS ET ACP VERTICALE

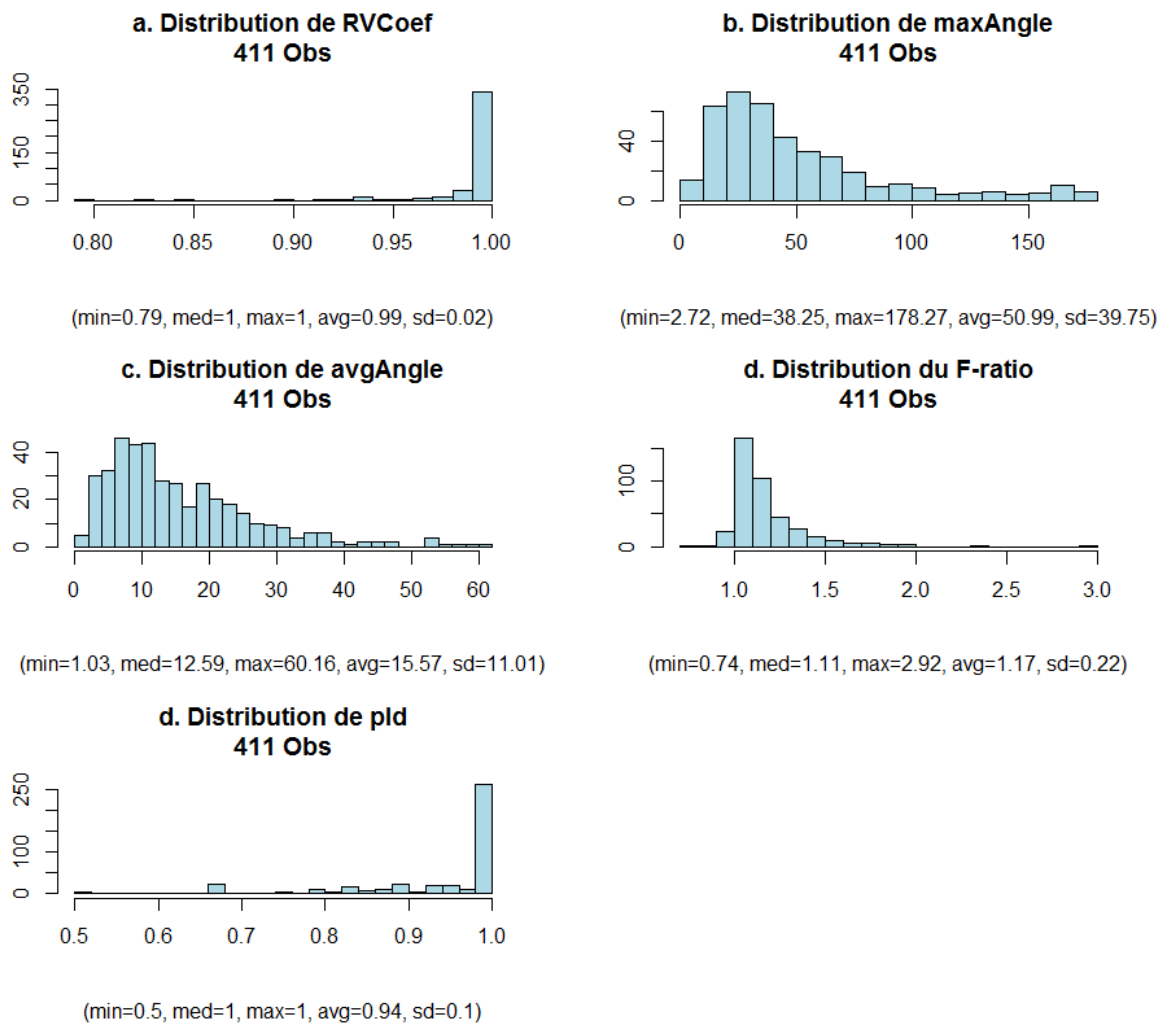


FIGURE A. 5 : RESULTATS DE LA COMPARAISON DE L'ACP ET L'ACP VERTICALE

	Méthode	Min	1 st Q.	Med	Mean	3 rd Q.	Max
Beta	ACP	1.008	1.394	1.785	1.905	2.226	5.108
Beta	ACP Verticale	1.028	3.634	5.032	5.416	6.705	20.180
N	ACP	1	1	2	1.876	2	5
N	ACP Verticale	1	4	6	6.363	8	24

TABLEAU A. 8 : DISTRIBUTIONS DES CRITERES DE DIMENSIONALITE DANS LA COMPARAISON DE L'ACP ET L'ACP VERTICALE

	Méthode	Min	1 st Q.	Med	Mean	3 rd Q.	Max
Beta	ACP	1.070	1.235	1.756	1.998	2.398	4.913
Beta	ACP Verticale	1.406	2.510	3.358	3.975	5.477	7.477
N	ACP	1	1	2	2	2	4
N	ACP Verticale	1	3	3	3.95	5	7

TABLEAU A. 9 : DISTRIBUTIONS DES CRITERES DE DIMENSIONALITE POUR LES JEUX DE DONNEES AVEC UN NOMBRE DE PRODUITS SUPERIEUR AU NOMBRE DE DESCRIPTEURS SIGNIFICATIFS UNIQUEMENT (49 JEUX DE DONNEES)

V. INTERPRETATION ET CONCLUSIONS

Ces différents résultats montrent que l'ACP et l'ACP Verticale sont généralement similaires. Cependant, de légers changements entre les directions des descripteurs de l'ACP et l'ACP verticale sont observés, et l'ACP semble donner une carte avec une meilleure discrimination que l'ACP Verticale (F-ratio, et diagnostic de paires de produits). On peut remarquer également que l'ACP Verticale est plus proche de l'ACP que l'ACP des corrélations et que l'ACP Verticale utilise plus de dimensions que l'ACP. Par conséquent, l'interprétation sensorielle semble plus riche pour l'ACP Verticale : l'hypothèse initiale est validée. Cependant, est-ce qu'on obtient réellement plus d'informations sur les différences entre les produits grâce à l'ACP Verticale ? Cette question mériterait d'être abordée dans des études ultérieures.

D. RAPPORT R4 : ETUDE DES PERFORMANCES DU PANEL SUR 177 JEUX DE DONNEES

I. PROBLEMATIQUE

Nous avons étudié ici différents aspects des performances unidimensionnelles du panel :

- distribution des indicateurs de performances (uni et multidimensionnels) du panel,
- lien avec la modalité sensorielle du descripteur,
- liens entre les différentes performances.

Par ailleurs, nous avons comparé le modèle MAM et le modèle à 2 facteurs, afin de quantifier l'apport du MAM par rapport au modèle à deux facteurs.

II. PROTOCOLE

1. ANALYSES STATISTIQUES

- Etude de la distribution des indicateurs de performances du panel

Les indicateurs de la table de performances sont ceux de la table MAM-CAP (Peltier, Brockhoff, Visalli, & Schlich, 2014). Des histogrammes du pourcentage d'attributs significatifs pour chaque effet par jeu de données ont été produits pour chaque performance (discrimination, accord, scaling, répétabilité). Les histogrammes des coefficients de scaling, niveau (hauteur des notes) et CME ont également été tracés.

- Les performances du panel ont ensuite été déclinées par type de descripteur

Le pourcentage de descripteurs significatifs pour chacun des différents indicateurs pour chaque modalité sensorielle a été calculé.

- Liens entre les différentes performances

Les coefficients de corrélations de Spearman entre les F discrimination/désaccord/scaling et répétabilité (CME) au niveau du panel ont été calculés.

- Comparaison du modèle MAM et du modèle à 2 facteurs

Les pourcentages moyens de descripteurs significatifs pour les effets produit (discrimination) et interaction (désaccord) (et scaling) ont été calculés pour le MAM et pour le modèle classique.

2. SELECTION DES JEUX DE DONNEES

Nous avons sélectionné des jeux de données équilibrés avec des descripteurs notés sur des échelles continues, avec plus de 4 sujets, plus de 3 produits et entre 2 et 100 descripteurs.

III. BASE DE DONNEES OBTENUE (177 JEUX DE DONNEES)

La base de données obtenue contient 177 jeux de données dont les caractéristiques sont décrites dans le Tableau A. 10 et la Figure A. 6.

Nombre de sujets	Nombre de produits	Nombre d'attributs	Nombre d'observations
1 590	1 007	3 741	617 211

TABEAU A. 10 : CARACTERISTIQUES DE LA BASE DE DONNEES SELECTIONNEE POUR L'ANALYSE DES PERFORMANCES DU PANEL

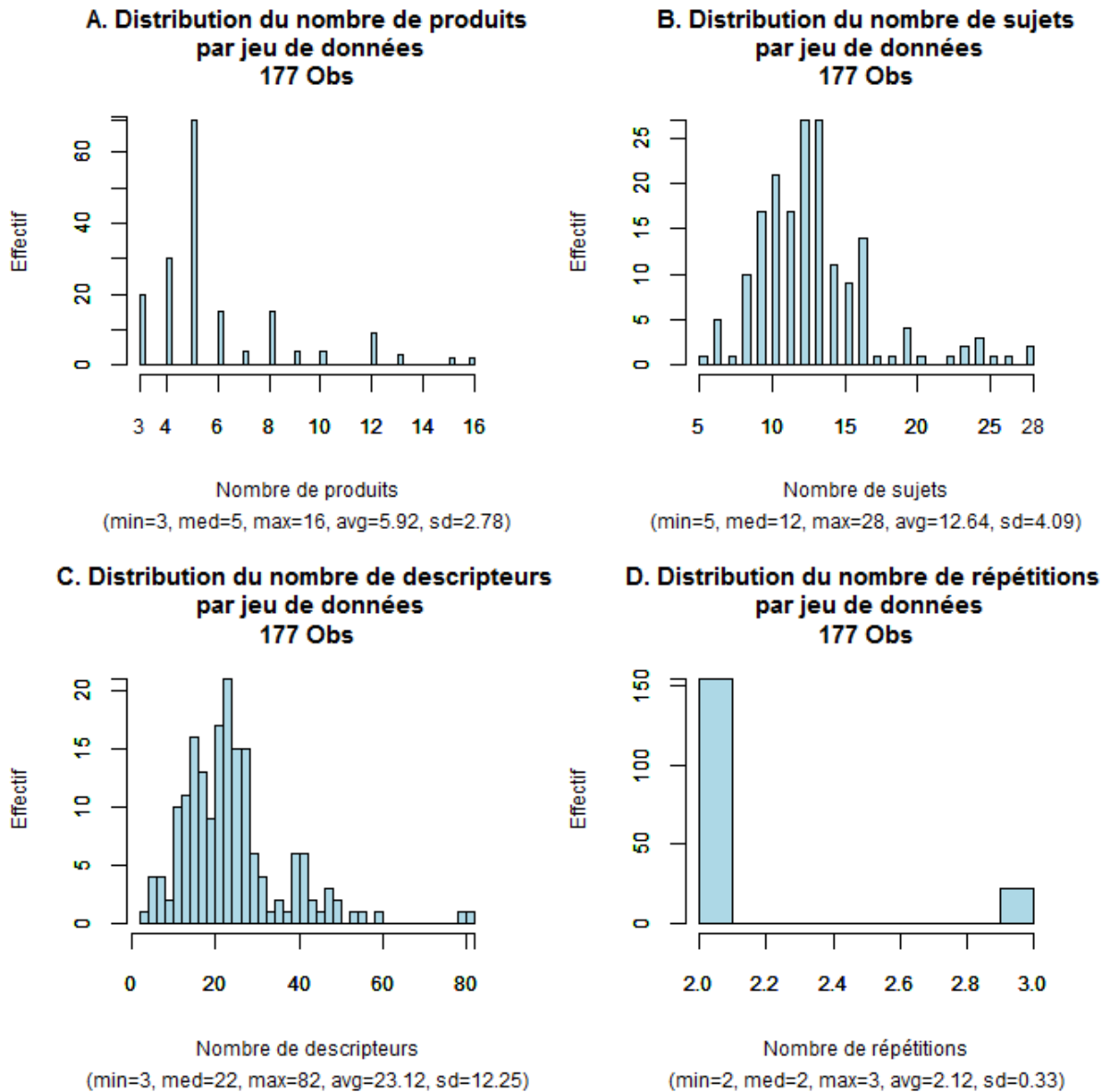


FIGURE A. 6 : DISTRIBUTIONS DES NOMBRES DE PRODUITS (A) SUJETS (B), DESCRIPTEURS (C) ET REPETITIONS (D) DES JEUX DE DONNEES ETUDIES POUR L'ANALYSE DES PERFORMANCES DU PANEL

IV. RESULTATS DES META-ANALYSES

1. DISTRIBUTION DES INDICATEURS DE PERFORMANCES

Les histogrammes ci-dessous représentent les distributions des différents indicateurs de performances (Figure A. 7).

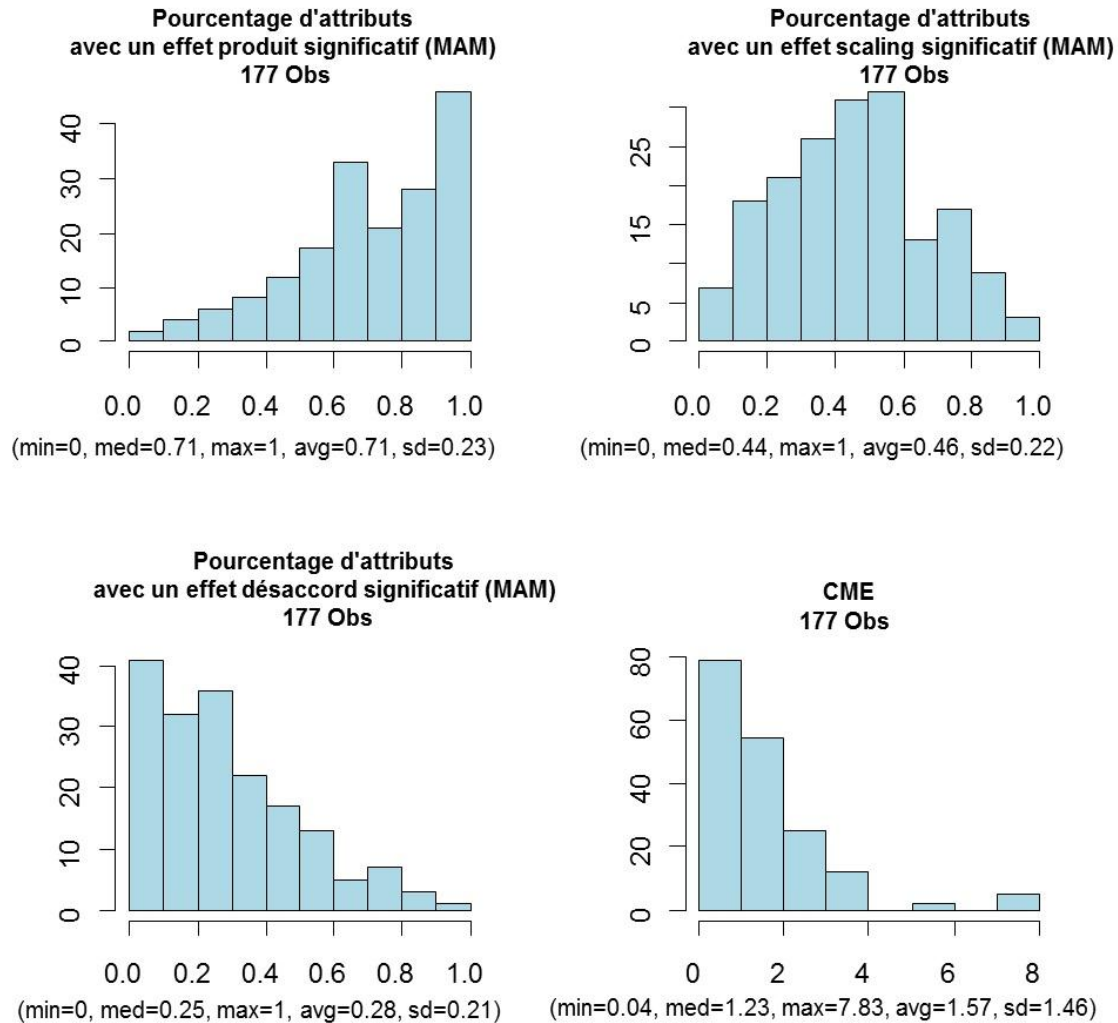


FIGURE A. 7 : DISTRIBUTIONS DES PERFORMANCES DU PANEL

2. RESULTATS SUR LES PERFORMANCES ET MODALITES SENSORIELLES ET INTERPRETATION

Le Tableau A. 11 résume les performances déclinées par modalité sensorielle. Les colonnes Discrimination, Accord et Scaling contiennent le pourcentage de descripteurs significatifs pour ces effets. La colonne Répétabilité correspond aux \sqrt{CME} moyens, et la colonne Niveau correspond à la moyenne des notes attribuées.

	N	Discrimination	Accord	Scaling	Répétabilité	Niveau
Apparence	348	84%	22%	46%	1.14	3.86
Odeur	407	57%	28%	41%	1.27	2.32
Saveur	872	69%	30%	43%	0.86	2.62
Texture	801	83%	29%	46%	1.10	3.61
Arôme	1369	68%	34%	51%	1.04	2.12
Arrière-gout	142	69%	37%	54%	0.99	2.15
Trigeminal	126	75%	29%	50%	1.18	2.49

TABLEAU A. 11 : POURCENTAGE DE DESCRIPTEURS SIGNIFICATIFS PAR MODALITE SENSORIELLE ET PERFORMANCES

Les descripteurs visuels et de texture ont un fort taux de discrimination, et sont notés haut sur l'échelle. Les descripteurs d'odeur ont été les moins discriminés, et la répétabilité a été meilleure pour les descripteurs de saveur.

3. LIENS ENTRE LES DIFFERENTES PERFORMANCES : RESULTATS ET INTERPRETATION

L'indicateur de Répétabilité correspond aux \sqrt{CME} . On note donc ici que plus la valeur de l'indicateur est élevée, moins le panel est répétable. Le Tableau A. 12 montre que la répétabilité semble corrélée au niveau des notes (0.46). En d'autres termes, moins le panel note « haut », plus le panel est répétable.

PANEL	Niveau	Discrimination	Scaling	Désaccord	Répétabilité
Level		0.148	-0.26	0.07	0.46
Discrimination	0.148		0.49	-0.23	-0.34
Scaling	-0.26	0.49		-0.12	-0.03
Désaccord	0.07	-0.23	-0.12		-0.12
Répétabilité	0.46	-0.03	-0.34	-0.12	

TABLEAU A. 12 : CORRELATIONS DE SPEARMAN ENTRE LES DIFFERENTES PERFORMANCES

Plus le panel est discriminant, plus il est répétable et moins il y a d'hétérogénéité de scaling et moins il est en désaccord. Cela caractérise probablement des jeux de données plus «

4. MAM VS ANOVA A DEUX FACTEURS : RESULTATS ET INTERPRETATION

Le Tableau A. 13 montre que l'utilisation du MAM a été pertinente pour environ un descripteur sur deux (une hétérogénéité significative de scaling a été observée pour 45% des descripteurs).

	Discrimination	Scaling	Désaccord
ANOVA à deux facteurs	64.75%	-	48.24%
MAM	70.78%	45.64%	28.30%

TABLEAU A. 13 : MOYENNE DES POURCENTAGES DE DESCRIPTEURS SIGNIFICATIFS POUR LES DIFFERENTS EFFETS SUIVANT LE MODELE CHOISI

De plus, l'utilisation du MAM par rapport au modèle à 2 facteurs a permis de gagner 6 % de significativité de l'effet produit, et de perdre 20% de désaccord significatif.

E. RAPPORT R5 : PERFORMANCES DES SUJETS EN FONCTION DE LEUR CARACTERISTIQUES

I. PROBLEMATIQUE

La problématique de cette étude est d'observer en quelle mesure le genre, l'âge et le fait de fumer ont un impact sur les performances unidimensionnelles (discrimination, scaling, accord et répétabilité) des sujets.

II. PROTOCOLE

1. OUTILS STATISTIQUES

CRITERES DE PERFORMANCES

Afin de répondre à cette question, les critères établis par Pineau (2006) sur la discrimination, l'accord et la répétabilité ont été utilisés. Nous avons également ajouté un critère pour prendre en compte le scaling.

- Discrimination : $\sqrt{MS_{Prod}/(MS_{Prod} + MS_{Error})}$
- Accord : corrélation entre les notes du sujet et celles du reste du panel
- Répétabilité : $\sqrt{MS_{Error}}$
- Scaling : coefficient de scaling β transformé par la fonction f (permettant de « symétriser le coefficient de scaling (chapitre 4, A))
- Level : différence entre la moyenne du sujet et celle du panel pour un descripteur fixé

Ces critères ont été calculés pour chaque sujet au sein de chaque jeu de données.

Afin de déterminer si les caractéristiques du sujet (genre, âge, fait de fumer) avaient un effet sur ses performances, différentes analyses ont été effectuées.

ANALYSE 1. La première analyse consiste à calculer une ANOVA suivant le modèle à deux facteurs (jeu de données et caractéristique) avec interaction et effet du jeu de données aléatoire.

$$Performance = caractéristique + jeu\ de\ données + caractéristique * jeu\ de\ données$$

Cette ANOVA est pondérée par la représentativité des caractéristiques (Pineau (2006)) dans le jeu de données. En effet, si, par exemple, un panel est constitué de 11 femmes et d'1 homme, l'estimation de l'effet de la caractéristique (le genre, ici) peut être biaisée. La pondération permet de prendre en compte la répartition des effectifs selon les catégories et correspond à la valeur suivante :

$$w = \frac{\text{effectif minimum observé dans une catégorie}}{\text{effectif moyen par catégorie}}$$

Ainsi, le poids affecté à chaque panéliste d'un profil sera égal à 1 lorsque le nombre de sujets de cette étude est parfaitement équilibré entre ces catégories (5 hommes et 5 femmes par exemple).

Avantage : autant d'observations que de sujets*descripteurs.

Inconvénient : dispositif déséquilibré.

ANALYSE 2

La seconde analyse est également une ANOVA pondérée (suivant le même système de pondération), suivie d'un test *post-hoc* (Tukey) mais cette fois, on considère la moyenne par jeu de données (on aura 2 observations par jeu de données, une pour chaque modalité) :

$$\text{Performance} = \text{caractéristique} + \text{jeu de données}$$

Avantages : le modèle est plus simple, et les données sont équilibrées.

Inconvénient : moins d'observations.

2. SELECTION DES JEUX DE DONNEES

On a sélectionné des jeux de données équilibrés, utilisant des échelles continues avec au moins 2 répétitions, entre 3 et 20 produits, et plus de 5 sujets.

III. BASE DE DONNEES RESULTANTE (177 JEUX DE DONNEES)

177 jeux de données ont été obtenus avec les caractéristiques présentées Tableau A. 14 et Figure A. 8.

Nombre de sujets	Nombre de produits	Nombre d'attributs	Nombre d'observations
1590	1007	3741	617211

TABLEAU A. 14 : CARACTERISTIQUES DE LA BASE DE DONNEES POUR L'ANALYSE DES PERFORMANCES DES SUJETS

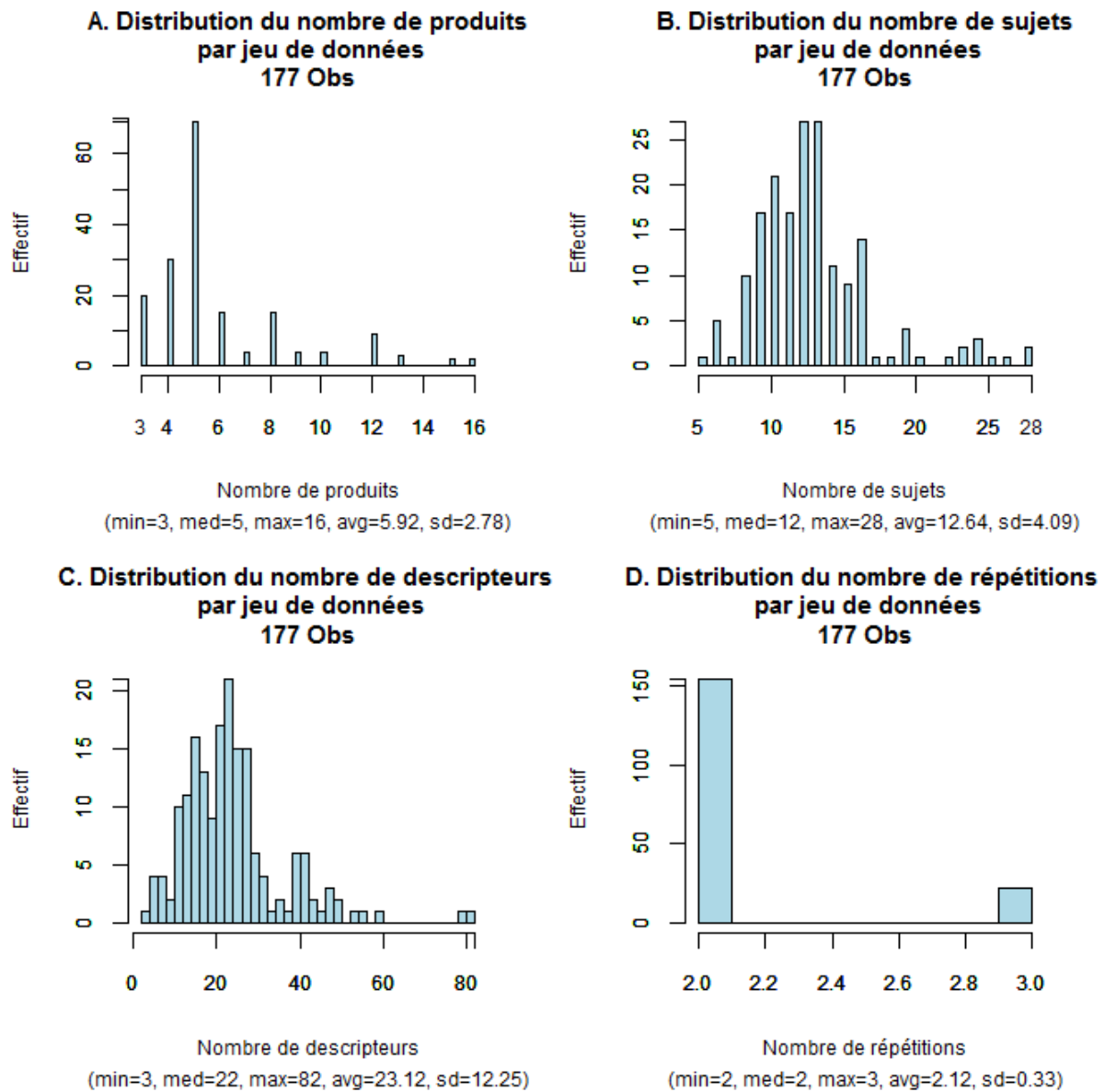


FIGURE A. 8 : DISTRIBUTIONS DES NOMBRES DE PRODUITS (A), SUJETS (B), DESCRIPTEURS (C) ET REPETITIONS (D) DES JEUX DE DONNEES OBTENUS POUR L'ANALYSE DES PERFORMANCES DES SUJETS

PERFORMANCES DES PANELISTES EN FONCTION DU GENRE

IV. RESULTATS DE LA META-ANALYSE ET INTERPRETATION

1244 sujets répartis dans 96 jeux de données ont été analysés. Les résultats de la méta-analyse sont résumés dans le Tableau A. 15.

	Analyse 1					Analyse 2				
	Stat F	Pvalue	Modalité	Groupes	n	Stat F	Pvalue	Modalités	Groupes	n
Discrimination	0.57	0.45	F	0.786 a	930	1.06	0.3	F	0.7907 a	96
			H	0.783 a	314			H	0.787 a	96
Désaccord	0.3	0.58	F	0.436 a	930	1.78	0.18	F	0.448 a	96
			H	0.432 a	314			H	0.437 a	96
Scaling	1.07	0.30	F	0.091 a	930	5.58	0.02	F	0.099 a	96
			H	0.039 a	314			H	0.000 b	96
Répétabilité	0.1	0.75	F	1.018 a	930	0.03	0.86	F	0.962 a	96
			H	1.077 b	314			M	0.966 a	96
Niveau	0.87	0.35	F	-0.034 a	930	2.60	0.11	F	-0.020 a	96
			H	0.069 b	314			H	0.039 a	96

TABLEAU A. 15 : RESULTATS SUR L'IMPACT DU GENRE SUR LES PERFORMANCES DES SUJETS

Les résultats montrent que les deux analyses s'accordent sur le fait qu'il n'existe pas d'effet du genre concernant la **discrimination**, l'**accord** et la **répétabilité** des panélistes. Cependant, si on devait dégager une tendance, les femmes seraient plus répétables que les hommes (valable pour les deux analyses). En ce qui concerne le **scaling**, on a obtenu la même tendance par les deux analyses mais la première n'a pas permis de mettre en évidence des effets significatifs de scaling, alors que la seconde a montré que **les hommes dispersent moins leurs notes que les femmes**. Finalement, les deux analyses ont montré une légère tendance à ce que les **hommes notent plus haut** que les femmes.

Pour conclure, le genre ne semble pas avoir un réel impact sur les performances des panélistes.

PERFORMANCES DES SUJETS EN FONCTION DU STATUT TABAGIQUE

855 sujets répartis dans 24 jeux de données ont été observés. Les résultats de la méta-analyse sont résumés dans le Tableau A. 15.

V. RESULTATS ET INTERPRETATION

	Analyse 1					Analyse 2				
	Stat F	Pvalue	Modalités	Groupes	n	Stat F	Pvalue	Modalités	Groupes	n
Discrimination	1.77	0.17	Non Fum.	0.801 b	605	3.50	0.038	Non Fum.	0.794 ab	24
			Ex Fum.	0.814 a	165			Ex Fum.	0.801 a	24
			Fumeur	0.762 c	85			Fumeur	0.781 b	24
Désaccord	1.04	0.35	Non Fum.	0.484 b	605	1.68	0.197	Non Fum.	0.472 a	24
			Ex Fum.	0.519 a	165			Ex Fum.	0.444 a	24
			Fumeur	0.367 c	85			Fumeur	0.437 a	24
Scaling	1.57	0.20	Non Fum.	-0.034 a	605	2.47	0.096	Non Fum.	-0.08 a	24
			Ex Fum.	-0.068 a	165			Ex Fum.	0.067 a	24
			Fumeur	-0.178 a	85			Fumeur	-0.218 a	24
Répétabilité	1.6	0.20	Non Fum.	0.899 b	605	1.07	0.35	Non Fum.	0.802 a	24
			Ex Fum.	0.80 a	165			Ex Fum.	0.850 a	24
			Fumeur	1.078 c	85			Fumeur	0.810 a	24
Niveau	0.15	0.85	Non Fum.	0.039 a	605	0.37	0.99	Non Fum.	0.039 a	24
			Ex Fum.	-0.069 b	165			Ex Fum.	0.066 a	24
			Fumeur	-0.041 ab	85			Fumeur	-0.023 a	24

TABLEAU A. 16 : EFFETS DU STATUT TABAGIQUE SUR LES PERFORMANCES DES SUJETS

Les deux analyses s'accordent sur le fait que les **fumeurs sont moins discriminants** que les autres panélistes, alors que les **anciens fumeurs semblent plus discriminants** que les autres panélistes. Cet effet est significatif pour l'analyse 2 ($p < 0.05$) et est juste une tendance pour l'analyse 1 ($p < 0.2$).

Les **fumeurs semblent moins disperser** leurs notes que les autres.

Les résultats n'ont pas montré d'effet du fait de fumer ou non sur l'**accord ou le niveau des notes**. Si on devait dégager une tendance, elle tendrait à affirmer que les fumeurs sont plus en désaccord et notent moins haut que les non-fumeurs. On ne peut pas conclure sur l'impact du fait de fumer ou non sur la répétabilité (résultats contradictoires).

PERFORMANCES DES SUJETS EN FONCTION DE L'ÂGE

VI. RESULTATS DE LA META-ANALYSE ET INTERPRETATION

1 719 sujets répartis dans 70 jeux de données ont été observés. Les résultats de la méta-analyse ont été résumés dans le Tableau A. 15.

	Analyse 1					Analyse 2				
	Stat F	Pvalue	Modalités	Groupe	n	Stat F	Pvalue	Modalités	Groupes	n
Discrimination	2.26	0.105	-40	0.776 a	328	2.20	0.11	-40	0.792 a	70
			40-60	0.775 a	843			40-60	0.789 a	70
			+60	0.770 a	548			+60	0.781 a	70
Désaccord	1.45	0.23	-40	0.391 a	328	4.93	0.008	-40	0.472 a	70
			40-60	0.427 b	843			40-60	0.442 b	70
			+60	0.463 c	548			+60	0.438 b	70
Scaling	20.31	<0.001	-40	-0.0001a	328	12.22	<0.001	-40	0.023 a	70
			40-60	0.078 a	843			40-60	0.06 a	70
			+60	-0.146 b	548			+60	-0.311 b	70
Repetabilité	21.26	<0.001	-40	1.253 a	328	18.8	<0.001	-40	1.009 a	70
			40-60	1.079 b	843			40-60	1.034 a	70
			+60	0.917 c	548			+60	0.840 b	70
Niveau	1.8	0.165	-40	-0.089 a	328	0.52	0.99	-40	-0.10 a	70
			40-60	0.028 b	843			40-60	0.051 a	70
			+60	0.002 b	548			+60	0.002 a	70

TABLEAU A. 17 : EFFETS DE L'ÂGE SUR LES PERFORMANCES DES SUJETS

Les deux analyses se sont accordées sur des effets significatifs de l'âge sur le scaling et la répétabilité : **les plus de 60 ans ont moins dispersé leurs notes** et sont clairement **plus répétables** que le reste du panel. De plus, on a observé le fait que les plus de 60 ans ont eu tendance ($p < 0.15$) à **moins discriminer leurs notes que les autres**. Enfin, les **moins de 40 ans** ont eu **tendance à noter moins haut** que le reste du panel.

Les deux analyses ont des résultats contradictoires sur l'accord, on ne peut donc pas conclure.

F. RAPPORT R6 : EQUIVALENCE ENTRE NOMBRE DE SUJETS SUPPLEMENTAIRES ET AJOUT D'UNE REPETITION

I. PROBLEMATIQUE

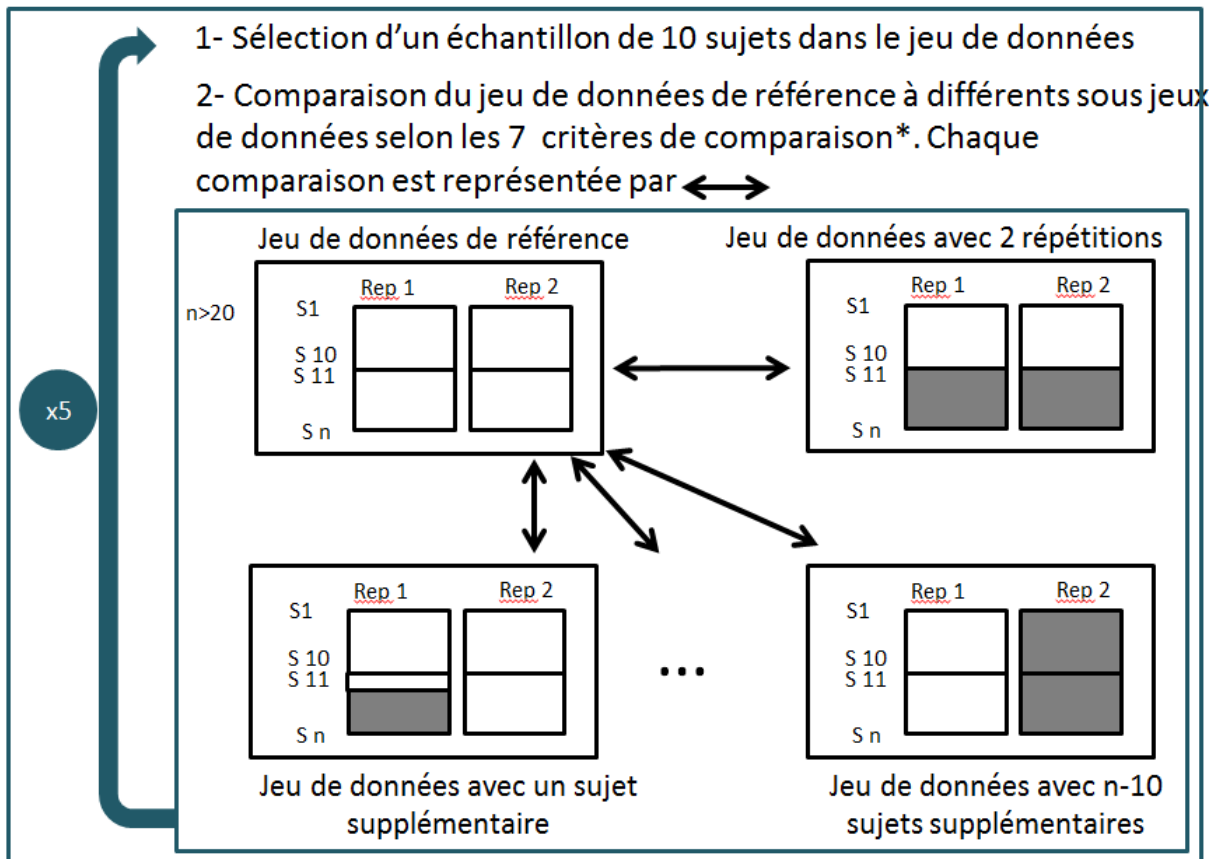
La question du nombre de sujets et du nombre de répétitions durant une expérimentation de profil sensoriel est cruciale pour tout animateur de panel. Augmenter le nombre de répétitions ou le nombre de sujets permet d'obtenir des résultats plus précis, mais a un coût financier plus élevé. L'animateur de panel doit donc trouver un compromis entre le niveau de précision des données et le coût de l'expérimentation. Dans ce contexte, la question de cette méta-analyse est la suivante : à combien de sujets supplémentaires équivaut une répétition supplémentaire (au sens de la qualité du jeu de données obtenu)?

II. PROTOCOLE

1. METHODES

UNE METHODE BASEE SUR L'EXTRACTION DE SOUS-JEUX DE DONNEES D'UN JEU DE DONNEES REFERENCE

La méthode adoptée lors de cette méta-analyse se base sur un jeu de données référence contenant un grand nombre de sujets et plusieurs répétitions. Pour chaque jeu de données, on effectue la démarche présentée Figure A. 9. Tout d'abord, on sélectionne aléatoirement dix sujets dans le jeu de données. Ce jeu de données constitue un « sous-jeu de données » à deux répétitions et 10 sujets (de S1 à S10). Ensuite, on restreint ce sous jeu de données à sa première répétition uniquement et on lui ajoute k sujets supplémentaires (k variant de 1 à n-10). L'ensemble de ces sous jeux de données est comparé au jeu de données de référence. De cette manière, on pourra observer le nombre de sujets équivalent à une répétition supplémentaire sur différents critères de comparaison.



* Discrimination (uni. et multi.), corrélation (uni. et multi.), groupes des produits(uni. et multi.), et comparaison de cartes

■ Représente des données qui ont été enlevées

FIGURE A. 9 : DEMARCHE PAR COMPARAISON DE SOUS-JEU DE DONNEES AU JEU DE DONNEES DE REFERENCE

L'objectif de cette étude est de trouver le nombre de sujets k pour lequel la « précision » du jeu de données à $(10+k)$ sujets sans répétition devient meilleure que celle à 10 sujets et deux répétitions.

Des critères ont été mis en place afin de comparer la discrimination, l'accord entre les notes, les groupes de produits obtenus (de manière uni et multidimensionnelle –par résultats de comparaison par paires-) et les cartes de chaque sous-jeu de données (jdd) au jeu de données de référence (ref). Ces critères ont été calculés de manière à être systématiquement compris entre 0 et 1, 1 indiquant une forte ressemblance entre les jeux de données.

CRITERES DE COMPARAISON ENTRE DEUX JEUX DE DONNEES

- Comparaison de la discrimination entre les jeux de données
 - **discrimination unidimensionnelle** : $1 - p_{ds}$ où p_{ds} est la proportion de descripteurs significatifs supplémentaires pour le jeu de données de référence ($p < 0.05$),
 - **discrimination multidimensionnelle** : rapport des F de la MANOVA $F_{Jdd} / F_{Réf}$.
- Comparaison des notes attribuées dans les jeux de données
 - **corrélations unidimensionnelles** : coefficient de corrélation moyen entre les descripteurs,

- **corrélations multidimensionnelle** : coefficient RV entre les deux tableaux des moyennes des produits obtenues.
- Comparaison des configurations produit obtenues sur les cartes de CVA (**rvMap**).
- Comparaison des groupes de produits obtenus lors des tests de comparaisons
 - **Groupes unidimensionnels** : les tests de Tuckey ont été effectués pour un seuil de 5%. Pour chaque paire de produits, un diagnostic « similaires » ou « différents » a ainsi été obtenu. La proportion de diagnostics identiques pour les deux jeux de données est calculée,
 - **Groupes multidimensionnels** : la proportion de paires de produits au diagnostic multidimensionnel équivalent (paire de produits similaires ou différentes) est calculée à l'aide du test de T2 de Hotelling.

On remarque que les indicateurs de discrimination peuvent éventuellement avoir une valeur supérieure à 1 (jeu de données de référence moins discriminant que le jeu de données extrait). Dans ce cas, on considèrera que l'indicateur en question est égal à 1.

Au total on a donc mis en place 7 critères de comparaison.

Pour chaque jeu de données référence, ce travail a été effectué sur 5 sous-jeux de données (représentant 5 échantillonnages différents de dix sujets).

ANALYSE DES CRITERES DE COMPARAISON OBTENUS

L'analyse des résultats obtenus peut être effectuée **en considérant chaque critère individuellement, ou en considérant l'ensemble des critères simultanément.**

REPRESENTATION GRAPHIQUE DES CRITERES CONSIDERES INDIVIDUELLEMENT

Pour un critère fixé, le nombre de sujets nécessaires pour atteindre le niveau d'une répétition supplémentaire a été calculé pour chacun des sous-jeux de données. Une distribution du nombre de sujets équivalents à l'ajout d'une répétition a ainsi été obtenue, dont on considère la moyenne et l'écart-type.

Présentons maintenant comment « le nombre de sujets équivalent à une répétition supplémentaire a été obtenu.

Les différents critères de comparaison peuvent être représentés par un graphique pour chacun des 13 jeux de données (Figure A. 10). En abscisses figure le nombre de sujets ajoutés, et en ordonnées la valeur du critère obtenu (1 si le jeu de données est similaire au jeu de données de référence). On place donc autant de points qu'on a ajouté de sujets dans le jeu de données, puis on les relie pour obtenir une courbe. Par ailleurs, on trace la droite horizontale correspond à la valeur de ce critère pour le jeu de données à 2 répétitions. Ce travail sera fait pour chaque tirage aléatoire de 10 sujets auquel sera associé une couleur spécifique (donc on aura 5 couleurs).

L'abscisse correspondant à l'intersection d'une courbe et d'une droite de la même couleur correspond donc au **nombre de sujets équivalent à l'ajout d'une répétition** (comme indiqué sur la Figure A. 10). Pour des raisons de place et de lisibilité, tous les graphiques ne sont pas présentés ici, et seul l'exemple des corrélations unidimensionnelles sera analysé.

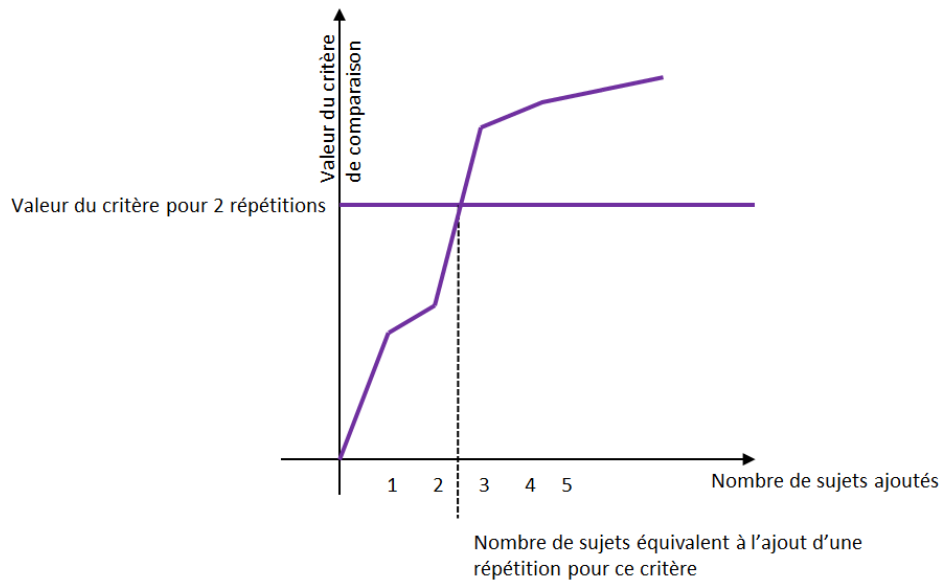


FIGURE A. 10 : REPRESENTATION GRAPHIQUE DE LA VALEUR DU NOMBRE DE SUJETS EQUIVALENT A L'AJOUT D'UNE REPETITION POUR UN CRITERE DE COMPARAISON FIXE ET UN ECHANTILLONNAGE FIXE.

ANALYSE DES CRITERES UTILISES SIMULTANEMENT

En considérant tous les critères simultanément, on a retenu le nombre de sujets k dans différentes situations (Figure A. 11) lorsque :

- Un des critères de différences pour k sujets supplémentaires devient plus grand (c'est à dire meilleur) que les critères de différences de référence (ce critère indique la 1^{ère} similarité, noté S_0).
- La moitié des critères de différences pour k sujets supplémentaires devient plus grand que les critères de différences de référence (noté $S_{1/2}$).
- Les deux tiers des critères de différences pour k sujets supplémentaires deviennent plus grand que les critères de différences de référence ($S_{2/3}$).
- Les trois quarts des critères de différences pour k sujets supplémentaires deviennent plus grand que les critères de différences de référence ($S_{3/4}$).
- Tous les critères de différences pour k sujets supplémentaires deviennent meilleurs que les critères de différences de référence (S_1).

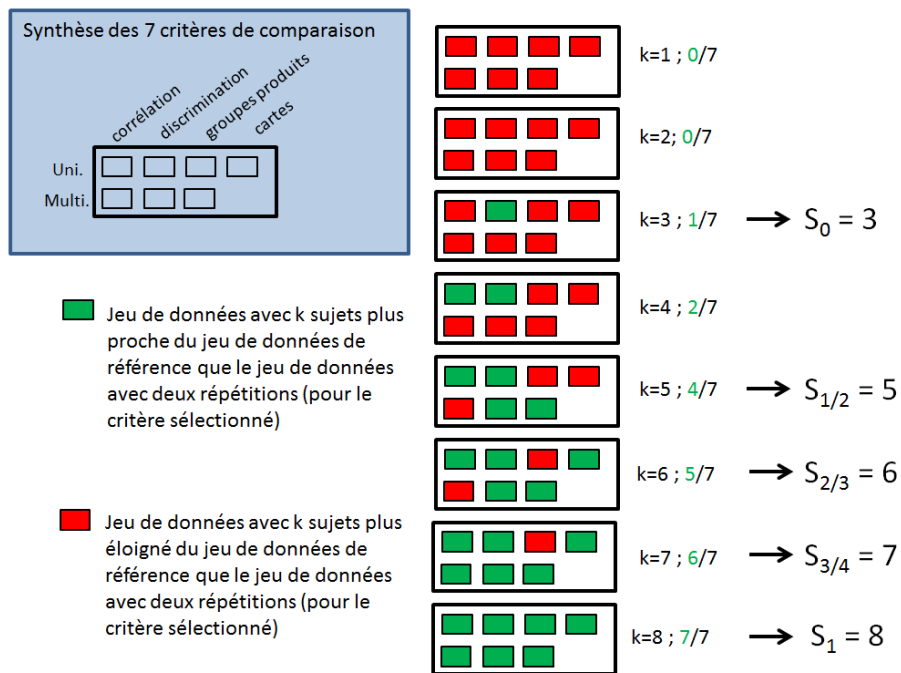


FIGURE A. 11 : SELECTION DU NOMBRE DE SUJETS EQUIVALENT A L'AJOUT D'UNE REPETITION PAR $S_0, S_{1/2}, S_{2/3}, S_{3/4}$ ET S_1

$S_0, S_{1/2}, S_{2/3}, S_{3/4}$ et S_1 ont également été calculés pour des jeux de données dont le nombre de sujets initial variait de 8 à 12 (précédemment, on s'était limité à 10). Ils ont également été calculés pour des jeux de données à 3 répétitions.

2. JEUX DE DONNEES

Nous avons sélectionné des jeux de données équilibrés, contenant plus de 20 sujets, entre 2 et 20 produits, moins de 100 descripteurs et au moins deux répétitions. **13 jeux de données** ont été obtenus et constituent les « jeux de données références ».

Pour l'étude à trois répétitions, après une pré-étude qui avait permis de montrer que le nombre de sujets à ajouter était moindre pour passer de deux répétitions à trois, nous avons conservé 5 jeux de données à 3 répétitions et moins de 15 sujets.

III. RESULTATS

1. ANALYSE EN CONSIDERANT LES CRITERES DE COMPARAISON INDIVIDUELLEMENT

La distribution du nombre de sujets supplémentaires équivalents à une répétition a indiqué un nombre minimal moyen de 7 sujets à ajouter relativement stable en fonction des critères, mais avec un écart type de plus de 5 (Tableau A. 18).

	Moyenne	Ecart type
Corrélation unidimensionnelle	7.9	5.29
Corrélation multidimensionnelle	7.32	5.8
Discrimination unidimensionnelle	7.1	5.4
Discrimination multidimensionnelle	5.1	5.7
Rv Map	4.8	5.4
Groupes unidimensionnels	8.3	5.3
Groupes multidimensionnels	8.5	6.8

TABLEAU A. 18 : MOYENNE ET ECART TYPE DU NOMBRE DE SUJETS MINIMAL EQUIVALENT A AJOUTER A UNE REPETITION

Le fait que l'écart-type soit aussi grand indique bien **de grandes variations entre les nombres de sujets nécessaires selon le jeu de données.**

LIMITES DE L'APPROCHE A L'AIDE DE L'EXEMPLE DU CRITERE DE CORRELATION UNIDIMENSIONNELLE

La Figure A. 12 permet de visualiser différents cas de figures selon l'échantillonnage initial (couleur de la courbe) et le jeu de données (graphique choisi). Pour commenter ces résultats, on a numéroté les jeux de données de 1 à 13 dans le sens de leur lecture.

Pour commencer, on peut remarquer que la plupart des courbes ont des allures croissantes : la ressemblance du jeu de données avec le jeu de données « référence » a bien augmenté avec le nombre de sujets. Cependant, ce n'est pas le cas pour toutes les courbes : par exemple, pour les jeux de données 7 ou 8, on peut voir que les courbes ne sont pas croissantes. Cela signifie probablement qu'il y a eu trop de différences entre les répétitions 1 et 2, peu importe les sujets choisis.

Ensuite, certaines courbes sont disjointes de la droite « référence ». Soit elles sont en dessous (courbe violette dans le jeu de données 13) - indiquant que le nombre maximal de sujets était encore insuffisant pour atteindre la précision d'une répétition supplémentaire - soit elles sont au-dessus (cas non présent ici, mais présent pour d'autres indicateurs), indiquant que le jeu de données à deux répétitions est plus éloigné du jeu de données référence que tous les jeux de données à une répétition ! Dans le premier cas, le nombre de sujets retenu correspond au nombre de sujets maximal que l'on peut ajouter dans le jeu de données de référence, dans le second, le nombre de sujets retenu est 0.

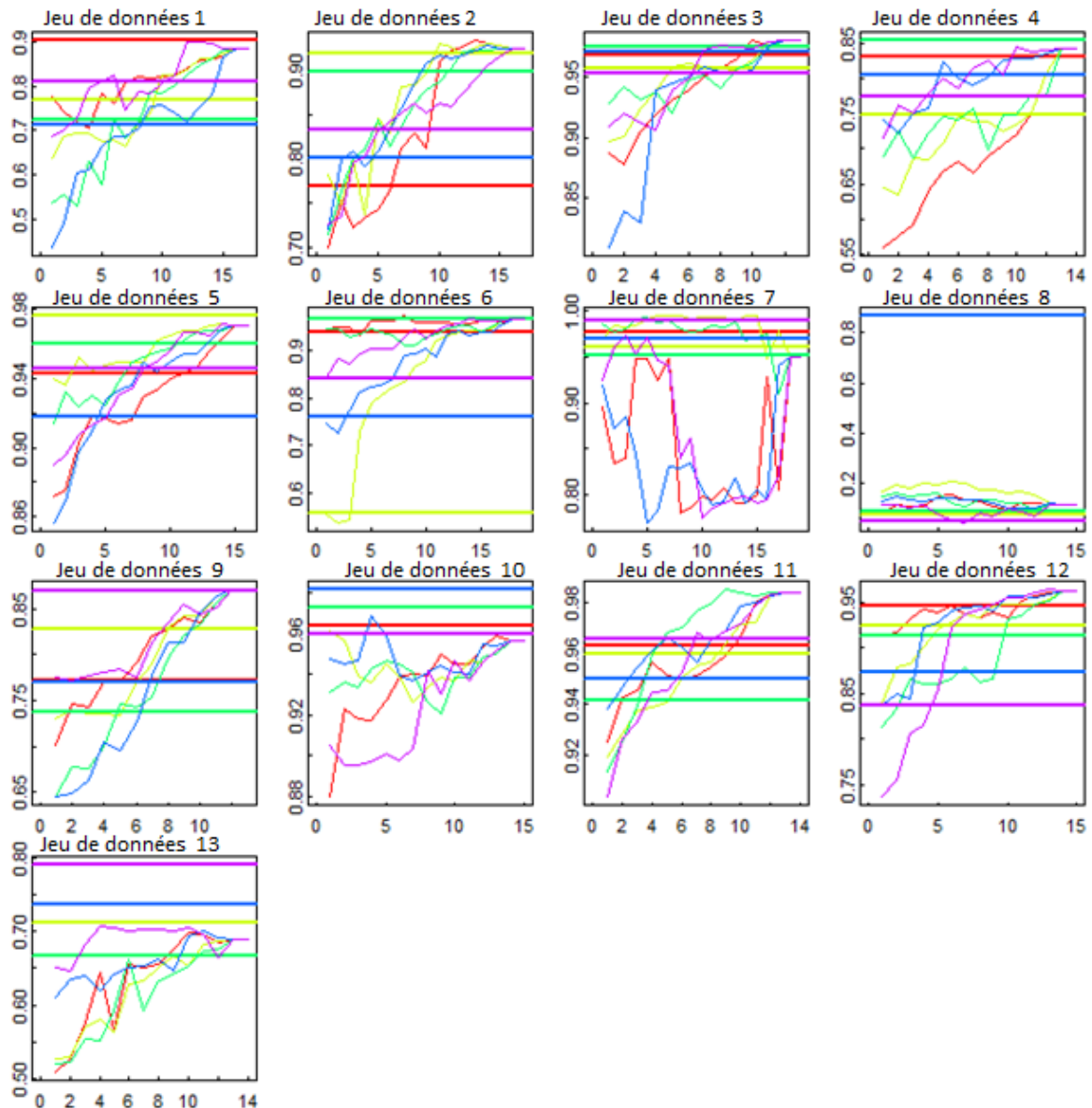


FIGURE A. 12 : VALEUR DE L'INDICATEUR DE CORRELATION UNIDIMENSIONNELLE EN FONCTION DU NOMBRE DE SUJETS AJOUTES POUR LES 13 JEUX DE DONNEES ET LES 5 ECHANTILLONNAGES.

Enfin, on a parfois observé de grandes différences selon l'échantillonnage choisi initialement : par exemple, le jeu de données 8 possède un critère à plus de 0.8 pour l'un de ces échantillonnage et 0.2 pour les autres.

L'ensemble de tous les critères a été étudié à l'aide de ce type de graphique, et on a retrouvé les mêmes cas de figures dans tous les graphiques. Cela montre la difficulté de quantifier le nombre de sujets équivalents à une répétition supplémentaire.

2. ANALYSE EN CONSIDERANT L'ENSEMBLE DES CRITERES DE COMPARAISON SIMULTANEMENT

Le tableau suivant représente la moyenne de $S_0, S_{1/2}, S_{2/3}, S_{3/4}$ et S_1 . n représente le nombre de sujets maximal à ajouter dans l'étude (soit le nombre de sujets du jeu de données de référence moins le nombre de sujets initial)

Nombre de sujets initial	S ₀	S _{1/2}	S _{2/3}	S _{3/4}	S ₁	n
8	0.05	2.88	7.42	10.48	15.08	16.64
9	0.4	5.8	9.5	11.97	14.94	15.64
10	0.2	4.2	8.5	11.05	14.11	14.64
11	0.21	4.2	8.9	11.23	13.2	13.64
12	0.5	5.4	9.1	10.98	12.03	12.64

TABLEAU A. 19 : NOMBRE MOYEN DE SUJETS SUPPLEMENTAIRES EQUIVALENT A L'AJOUT D'UNE DEUXIEME REPETITION SELON LA TAILLE INITIALE DU PANEL

On remarque assez peu de changement suivant le nombre de sujets contenu dans le panel. Le nombre de sujets à ajouter est d'environ 4 pour que la moitié des critères de différences soit meilleurs pour le jeu de données à sujets ajoutés ; 8 pour les deux tiers et 11 pour les trois quarts.

Le même travail a été effectué pour passer de deux à trois répétitions.

	S ₀	S _{1/2}	S _{2/3}	S _{3/4}	S ₁	n
8	0	2.16	4.2	4.6	6.96	12
9	0.07	2.12	4.08	5.08	7.04	11
10	0.1	3.16	4.4	5.8	7.16	10
11	0.1	2.36	3.12	5.12	7.6	9
12	0	2.36	2.88	3.8	5.48	8

TABLEAU A. 20 : NOMBRE DE SUJETS A AJOUTER DANS L'ETUDE EQUIVALENT A L'AJOUT D'UNE TROISIEME REPETITION EN FONCTION DU NOMBRE DE SUJETS INITIAL ET DES DIFFERENTS INDICATEURS

L'ajout d'une troisième répétition semble être équivalent à ajouter 2 à 6 sujets maximum dans le panel.

IV. INTERPRETATION

Le nombre moyen de sujets supplémentaires correspondant à l'ajout d'une deuxième répétition vaut entre 7 et 11 sujets (si on considère à la fois les résultats unidimensionnels et les résultats unidimensionnels pour S_{2/3} et S_{3/4}). Si on considère qu'il suffit que la moitié des critères de comparaison doivent être vérifiés, ce nombre est réduit à 4 à 6 sujets. L'ajout d'une troisième répétition est quant à elle équivalente à l'ajout de 4 à 5 (ou 2 ou 3 si on considère S_{1/2}) sujets supplémentaires. Cependant, on a pu observer une grande variabilité du nombre de sujets équivalents à une répétition supplémentaire.

Gower, J. L., S.; Le Roux, N., . (2011). Understanding biplot. In L. John Wiley and Sons.

Peltier, C., Visalli, M., & Schlich, P. (2015a). Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies. *Food Quality and Preference*, 40, Part B(0), 326-333.

Peltier, C., Visalli, M., & Schlich, P. (2015b). Taking the scaling effect into account in multidimensional analysis with MultiMAM. *Journal of Statistical Software*, (to be submitted).

Pineau, N. (2006). *La performance en analyse sensorielle, une approche base de données*.

Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlation. In T. Naes & E. Risvik, *Multivariate analysis of data in sensory science*.

- Dettmar, B., Plattner, M., & Vetter, C. (2014). Underdetermination in Multivariate data analysis. In, *13th symposium on statistical methods for the food industry*.
- Gower, J. L., S.;Le Roux, N., . (2011). Understanding biplot. In L. John Wiley and Sons.
- Peltier, C., Brockhoff, P. B., Visalli, M., & Schlich, P. (2014). The MAM-CAP table: A new tool for monitoring panel performances. *Food Quality and Preference*, 32, Part A(0), 24-27.
- Peltier, C., Visalli, M., & Schlich, P. (2015a). Comparison of Canonical Variate Analysis and Principal Component Analysis on 422 descriptive sensory studies. *Food Quality and Preference*, 40, Part B(0), 326-333.
- Peltier, C., Visalli, M., & Schlich, P. (2015b). Taking the scaling effect into account in multidimensional analysis with MultiMAM. *Journal of Statistical Software*, (to be submitted).
- Pineau, N. (2006). *La performance en analyse sensorielle, une approche base de données*.
- Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlation. In T. Naes & E. Risvik, *Multivariate analysis of data in sensory science*.

ANNEXE 3 :

AIDE DES PACKAGES TIMESENS

A. AIDE DU PACKAGE CVA

R Documentation

CVApack-package {CVApack}

CVApack-package

Description

Allows plotting Canonical Variate Analysis

Details

Package: CVA

Type: Package

Version: 1.0

Date: 2014-11-03

License:

To obtain CVA results, use the [CVA](#) function on your dataset. Then, use the [PlotCVA](#) function to plot the corresponding graph.

Author(s)

Caroline Peltier

Maintainer: <caroline.peltier@dijon.inra.fr>

Examples

```
data(cheeses)
```

```
PlotCVA(CVA(cheeses),representation="biplot")
```

[Package *CVApack* version 1.0 [Index](#)]

CVA

Description

Performs the Canonical Variate Analysis (CVA) for a dataset of sensory profile

Usage

```
CVA(extendedData, test = "Hotelling-Lawley", option = "Two-Way ANOVA", hotellingTableBool = TRUE, nbAxes = "auto", alpha = 0.1, productName = "ProductCode", subjectName = "SubjectCode", replicateName = "Replicate", sessionName = "Session", representation = "biplot")
```

Arguments

extendedData	data whose colnames are productName, subjectName, replicateName, sessionName, first Attribute, , last Attribute
test	MANOVA statistic used for the test. By default "Hotelling-Lawley" but "Wilks", "Roy" and "Pilai" are also available
option	"Two-Way ANOVA" or "One-Way ANOVA". This corresponds to the choice of the model used in CVA. By default "Two-way ANOVA". Note that the subject effect is random.
hotellingTableBool	if TRUE, the Hotelling table (where the multidimensional differences between products are tested) is calculated
nbAxes	either a number of axes or "auto" which return a number of axes equaling the number of significant dimensions. "auto" by default
alpha	limit of the test of significant dimensions. By default 0.05
productName	name of the column containing the products. By default "ProductCode"
subjectName	name of the column containing the subjects. By default "SubjectCode"
replicateName	name of the column containing the replicates. By default "Replicate"
sessionName	name of the column containing the sessions. By default "Session"
representation	if "biplot", the graph resulting of this CVA is a biplot, ifelse, it is a graph composed with a product map and an attribute map

Value

A list containing :

IndivCoord coordinates of the products

VarCoord	coordinates of the attributes
NbDimSig	number of significant dimensions
HotellingTable	table containing the results of T2 Hotelling tests for all product pairs
ConditioningOfW	condition index of W. If ConditioningOfW >30, the matrix is ill-conditioned and CVA can be unstable
B	covariance matrix of product effect
W	covariance matrix of interaction
EigenVectors	W-orthonormalized eigenvectors of $W^{-1}B$
EigenValues	eigenvalues of $W^{-1}B$
Stats	list containing the MANOVA statistics (statistics, approximated F and pvalue)
decomposition	table with the decomposition of each score according to the different effects
CenteredProductSubjectTable	centered product by subject table
nbAxes	number of axes, useful if nbAxes="auto" was chosen
wDemi	matrix of weights used in the biplot
option	returns the chosen model
representation	if "biplot", the plot related to this CVA will be a biplot. If else, it will be two graphs: a product map and an attribute map

Author(s)

Caroline Peltier

See Also

[PlotCVA](#)

Examples

data(cheeses)

resCVA=CVA(cheeses)

[Package *CVApack* version 1.0 [Index](#)]

PlotCVA

Description

Plots a CVA graph

Usage

```
PlotCVA(res.CVA, panellists = "none", alpha = 0.1, conflnt = 0.9, ellipsesType = "barycentric",  
productColors = NULL, returnX = FALSE, returnY = FALSE, linkBetweenProducts = TRUE,  
supplIndividualsToPlot = FALSE, type = "R", fileName = "CVA", mainTitle = "CVA")
```

Arguments

res.CVA	result of the function CVA
panellists	representation of the panellists on the map: "none" if the panellist are not represented on the map, "labels" if they are represented by their names, or "points" if they are represented by points
conflnt	limit for the confidence ellipses. By default 0.9
ellipsesType	"barycentric" or "individual". Barycentric if an ellipse represents a zone of confidence where the mean point is (with a probability of conflnt), "individual" if an ellipse represents a zone where the subjects scores are (with a probability of conflnt)
productColors	vector of colors for products
returnX	if TRUE, the X-axis is reversed. FALSE by default
returnY	if TRUE, the Y-axis is reversed. FALSE by default
linkBetweenProducts	if TRUE, two products which were equal with Hotelling T2 test are linked by a segment. TRUE by default
type	type of the output. "R" corresponds to a classical R-graph, whereas "png", "jpg" and "wmf" create a file in the working directory.
fileName	name of the file containing the CVA in the working directory
mainTitle	title of the map

Author(s)

Caroline Peltier

See Also

[CVA](#)

Examples

```
resCVA=CVA(cheeses,representation="biplot")
```

```
PlotCVA(resCVA)
```

```
resCVA=CVA(cheeses,representation="twoMaps")
```

```
PlotCVA(resCVA)
```

[Package *CVApack* version 1.0 [Index](#)]

B. AIDE DU PACKAGE MAMCAP

MAMCAP-package {MAMCAP}

R Documentation

MAMCAP-package

Description

This package allows to use the Mixed Assessor Model for the panel and panelist performances in profile sensory studies (MAM-CAP table)

Details

Package: MAMCAP

Type: Package

Version: 1.0

Date: 2013-07-12

License: GPL (>= 2)

This package allows to plot performances table. CAPTable and ReadProfileData are the most important functions.

[ReadProfileData](#) allows to get a "Profile Object" required in CAPTable.

[CAPTable](#) produces CAP table (Schlich, www.sensobase.fr) or MAM-CAP Table (Peltier et al., 2013).

Thus, to get a CAPTable, there are three steps:

To have data whose colnames are: ProductCode, SubjectCode, AttributeCode, Replicate and Score in an extern file

To read this file with ReadProfileData: `myProfileObject=ReadProfileData("data.txt")`

To run CAPTable : `CAPTable(myProfileObject,model="TwoWayANOVA")`

To run MAM-CAPTable : `CAPTable(myProfileObject,model="MAM")`

Author(s)

Caroline Peltier, Per Brockhoff, Michel Visalli, Pascal Schlich

Maintainer: Caroline Peltier <caroline.peltier@dijon.inra.fr>

References

Peltier, C., Brockhoff, P., Visalli, M. Schlich, P., The MAM-CAP table a new tool for monitoring panel performances. (accepted in Food Quality and Preferences in July 2013)

Brockhoff P. , Schlich P.; Skovgaard I, Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. (accepted in Food Quality and Preferences in July 2014)

Examples

```
data(wine)
```

```
CAPTable(wine)
```

[Package **MAMCAP** version 1.0 [Index](#)]

CAPTable

Description

This function build an html file (CAPTable.html) containing a panel and panelist performance table (CAP or MAM-CAP table) in current working directory.

Usage

```
CAPTable(profileObject, model = "MAM", panelLimit = 0.05, indivLimit = 0.05, language = "en",  
output = "CAPTable", correction = FALSE,correlationTest="kendall")
```

Arguments

- profileObject** profile Object resulted from [ReadProfileData](#)
- model** character. "TwoWayANOVA" or "MAM": model of performance tests.
"TwoWayANOVA" refers to the first Schlich CAP table available on the SensoBase (www.sensobase.fr) and does not take the scaling into account.
"MAM" refers to the MAM-CAP table, using the performances tests of Mixed Assessor Model (2014)
- limitPanel** limit of the statistical tests concerning panel performances
- limitIndiv** limit of the statistical tests concerning individual performances. When the chosen model is CAP, limitIndiv is only the limit of discrimination tests (Kendall agreement test limit = 0.2 and repeatability test= 0.01)
- correlationTest** "none" or "kendall","pearson" or "spearman". If "none", the disagreement test of panelist consists in comparing the disagreement of one panelist to his repeatability. If one of the three correlations tests ("spearman","kendall" or "pearson") is chosen, the test consists in testing the correlations of the panelist scores with the panel scores ($p < 0.2$)
- language** character. "fr" for french and "en" for english. Language of the legend.
- output** the name of the html file is created in the current working directory
- correction** logical. If TRUE, limitIndiv is divided by the number of panelists

Author(s)

Caroline Peltier, Per Brockhoff, Michel Visalli, Pascal Schlich

References

www.sensobase.fr

Brockhoff P. , Schlich P.; Skovgaard I, Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. (accepted in Food Quality and Preferences in July 2013)

Peltier C., Brockhoff, P., Visalli, M., Schlich, P. The MAM-CAP table: a new tool to monitor panel performances. (accepted in Food Quality and Preferences in July 2013)

Examples

```
wine=ReadProfileData("http://www.timesens.com/docs/wines.txt",sep="\t")
```

```
CAPTable(wine,model="TwoWayANOVA")
```

or

```
data(wine)
```

```
CAPTable(wine)
```

```
CAPTable(wine,model="TwoWayANOVA")
```

[Package *MAMCAP* version 1.0 [Index](#)]